5-11-2009

Independent component analysis (ICA) applied to ultrasound image processing and tissue characterization

Di Lai

Independent Component Analysis (ICA) Applied to Ultrasound Image Processing and Tissue Characterization

by

Di Lai

B.S., Biomedical Electronic Engineering, Xi'an Jiaotong University, 1989
M.S., Imaging Science, Rochester Institute of Technology, 2000

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology

May. 11, 2009

Signature of the Author _____

Accepted by _____
Coordinator, Ph.D. Degree Program Date

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Di Lai
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation required for the
Ph.D. degree in Imaging Science

N.A.H.K.Rao, Ph.D, Dissertation Advisor

Maria Helguera, Ph.D

Eli Saber, Ph.D

Joseph G. Voelkel, Ph.D, Committee Chair

Date

DISSERTATION RELEASE PERMISSION
ROCHESTER INSTITUTE OF TECHNOLOGY
CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE

Title of Dissertation:
Independent Component Analysis (ICA) Applied to Ultrasound Image Processing and Tissue Characterization

I, Di Lai, hereby grant permission to Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature _____
Date

Acknowledgments

First of all, I would like to thank my advisor Dr. Navalgund Rao, for his invaluable guidance, wisdom, encouragement, inspiration, kindness and patience in the past ten years. I will never forget the numerous discussions that we had in his office, in the reading room and in the lab. His insight into some tricky problems always enlightened my mind when I was entangled. I am also indebted to him for all the help regarding my Ph.D study time extension and all the other difficulties that I encountered on this long and winding road. The most important thing I would like to thank him is that he made me know how big a pleasure scientific research is.

I would like to thank Dr. Maria Helguera for her precious advice, concerns and help in the past ten years. Without her help, I couldn't have finally passed the defense and finished this dissertation.

I would like to thank Dr. Eli Saber and Dr. Joseph Voelkel, for their generous efforts monitoring the quality of this doctoral research and countless informative advices on improving my presentation and writing skills. Without their strict requirements, I couldn't have finally finished this dissertation.

I would like to thank my former colleague Dr. Chung-hui Kuo at Eastman Kodak for his support and help. It was he who first introduced ICA to me in 2002. In the past years we had so many discussions and I always treated him as a "lifeline" when I had some difficult technical problems.

I would like to thank Dr. Anthony Vodacek for his kind help regarding my Ph.D study time extension. I also would like to thank Dr. Andrew Moore, the Dean of Graduate

Studies, for his kind approval of the extension application. I would also like to thank the Center for Imaging Science for its generous toleration of my unusual long term of study.

I would like to thank Mrs. Sue Chan for her help in the past years for things such as credit registration, inquiry, transfer and extension application.

I would like to thank my old friend Dr. Tim Barker in England, for his continuous encouragement on my Ph.D study.

I would like to thank my brother Hanxu for his endless concern and help during all these years.

I would like to thank my parents for giving me the spiritual strength for overcoming all the difficulties in all these years. Although my father can't see what I have done, I believe he will smile in the heaven and be proud of me. In the past three years, my mother has accompanied me to go through all the torturing times in life. Without her encouragement, I might have given up.

Finally, I would like thank my wife, Xia, for her entire devotion to the family, taking care of the children, tolerating the pressure of life and encouraging me to pursue my dream. I also thank my two little daughters, Olivia and Isabella for bringing me so much happiness in life.

Dedication

This dissertation is dedicated to my father, Shuji Qin, professor of Southwest University of China. May your soul rest in peace.

Also it is dedicated to my dear mother Yinbo Zhang, my beloved wife Xia Zhao, and my angle-like daughters Olivia and Isabella.

Abstract

As a complicated ubiquitous phenomenon encountered in ultrasound imaging, speckle can be treated as either annoying noise that needs to be reduced or the source from which diagnostic information can be extracted to reveal the underlying properties of tissue. In this study, the application of Independent Component Analysis (ICA), a relatively new statistical signal processing tool appeared in recent years, to both the speckle texture analysis and despeckling problems of B-mode ultrasound images was investigated.

It is believed that higher order statistics may provide extra information about the speckle texture beyond the information provided by first and second order statistics only. However, the higher order statistics of speckle texture is still not clearly understood and very difficult to model analytically. Any direct dealing with high order statistics is computationally forbidding. On the one hand, many conventional ultrasound speckle texture analysis algorithms use only first or second order statistics. On the other hand, many multichannel filtering approaches use pre-defined analytical filters which are not adaptive to the data. In this study, an ICA-based multichannel filtering texture analysis algorithm, which considers both higher order statistics and data adaptation, was proposed and tested on the numerically simulated homogenous speckle textures. The ICA filters were learned directly from the training images. Histogram regularization was conducted to make the speckle images quasi-stationary in the wide sense so as to be adaptive to an ICA algorithm. Both Principal Component Analysis (PCA) and a greedy algorithm were used to reduce the dimension of feature space. Finally, Support Vector Machines (SVM) with Radial Basis Function (RBF) kernel were chosen as the classifier for achieving best classification accuracy. Several representative conventional methods, including both low and high order statistics based methods, and both filtering and non-filtering methods, have been chosen for comparison study. The numerical experiments have shown that the proposed ICA-based algorithm in many cases outperforms other algorithms for comparison. Two-component texture segmentation experiments were conducted and the proposed algorithm showed strong capability of segmenting two visually very similar yet different texture regions with rather fuzzy boundaries

and almost the same mean and variance. Through simulating speckle with first order statistics approaching gradually to the Rayleigh model from different non-Rayleigh models, the experiments to some extent reveal how the behavior of higher order statistics changes with the underlying property of tissues. It has been demonstrated that when the speckle approaches the Rayleigh model, both the second and higher order statistics lose the texture differentiation capability. However, when the speckles tend to some non-Rayleigh models, methods based on higher order statistics show strong advantage over those solely based on first or second order statistics. The proposed algorithm may potentially find clinical application in the early detection of soft tissue disease, and also be helpful for better understanding ultrasound speckle phenomenon in the perspective of higher order statistics.

For the despeckling problem, an algorithm was proposed which adapted the ICA Sparse Code Shrinkage (ICA-SCS) method for the ultrasound B-mode image despeckling problem by applying an appropriate preprocessing step proposed by other researchers. The preprocessing step makes the speckle noise much closer to the real white Gaussian noise (WGN) hence more amenable to a denoising algorithm such as ICS-SCS that has been strictly designed for additive WGN. A discussion is given on how to obtain the noise-free training image samples in various ways. The experimental results have shown that the proposed method outperforms several classical methods chosen for comparison, including first or second order statistics based methods (such as Wiener filter) and multichannel filtering methods (such as wavelet shrinkage), in the capability of both speckle reduction and edge preservation.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivations and Methodology

1.1.1 Motivations

In recent years, a new statistical signal processing technology called Independent Component Analysis (ICA) has attracted people's attention. The detailed introduction of ICA will be given in chapter 2. Based on the literature review given in Section 1.7, it is clear that classical ultrasound texture analysis and tissue characterization methods are either based on lower order statistics (first or second order) or based on multichannel filtering using pre-defined filter bank. Since ICA is related to higher order statistics and ICA filter bank is data-adaptive, we are motivated to investigate how this new technology could be applied to ultrasound texture analysis and ultrasound tissue characterization (UTC) so as to improve the efficiency of of tissue classification, characterization and texture based segmentation. First, second and higher order statistics (FOS, SOS and HOS) will be introduced in Section 1.6. General introduction of ultrasound imaging and speckle will be given in Sections from 1.2 to 1.5.

In many situations first order statistics give us most of the information about the properties of tissues. Several 1-D probability distributions based on ultrasound physics models have made a lot of achievements in the past years. Several non-model based second order

statistics methods, without making use of any *a priori* assumptions about tissue histology but merely using methods developed in the general fields of image analysis, have also shown some success. However, a question one may ask is: if the first or second order statistical measures for differentiating two textures are not available or weak (for example, one cancer texture and one normal texture may have almost the same mean and variance values), or just simply can not be used as valid differentiation criteria (for example, when two textures are obtained by different ultrasound insonification energy level so that the contrast difference is largely due to instrumental factors), can we rely on the higher order statistics to tackle this challenge? Early detection of soft tissue disease used to be a difficult problem and is also a significant research topic. The slight pathological change in tissue histology may not drastically change the first or second order statistics of the texture but the visual pattern (via HOS) may provide us extra cues. Taking consideration of HOS for texture analysis is one of the major motivations that we resort to ICA.

Texture analysis by multichannel filtering is similar to how human visual system processes the texture information. Classical multichannel filtering methods such as wavelet and Gabor transform use pre-defined analytical filters to decompose a texture image. On the other hand, there are some evidences shown that human visual system tends to decompose a visual pattern into uncorrelated (or independent) components so as to effectively code the information in multichannel [74][83][64]. This process is more data-adaptive (hence more effective) than using pre-defined fixed filters. An experienced radiologist may be able to differentiate the slight changes in the textures which are related to changes of tissue properties and ignore the interference caused by the contrast change. The ICA-based texture analysis is such kind of method which is similar to human visual system. Taking consideration of the data-adaptation of filter bank is another motivation that we intend to use ICA.

Direct dealing with HOS will encounter great computational difficulty. Since ICA is a novel information redundancy reduction tool so that the computation burden can be greatly

reduced. Instead of using sophisticated ultrasound physics models which are difficult to obtain (even for SOS), the proposed ICA-based algorithm takes a general image processing approach and is "blind" in nature (that is, using none or little information of the tissue microstructure).

We are also interested in knowing how a higher order statistical method behaves when the underlying tissue properties change gradually. It is still not clear on this aspect in ultrasound imaging community (even for second order statistics). The change of tissue properties can be related to the ultrasound physics models in different ways (for example, the deviation from the Rayleigh model to different sub-Rayleigh models such as K and Rice distributions due to the changes of tissue properties such as scatterer number density and scatterer spatial randomness). By using the ICA-based approach as a dissecting tool, we may gain some insight into how high order statistics change with underlying tissue properties. This is another important motivation of our work. By correlating the gradually changing properties of simulated tissues with the feature vectors obtained from the texture images, further quantitative UTC algorithms are possible to be developed.

Since speckle texture analysis and speckle reduction are just like two sides of the same coin, we can also apply the same technology, ICA, to the despeckling task. The same motivations lie in the consideration of both HOS and data-adaptation.

1.1.2 Methodology

In order to demonstrate the advantage of the proposed ICA-based algorithm over other conventional algorithms, we have chosen several algorithms (including both lower order statistics methods and multichannel filtering methods using fixed bases) for comparative study. The details are given in Chapter 4.

In this study, simulated tissues and the corresponding B-mode speckle texture images are used. There are several reasons of using simulated images and some detailed discussion will be given in chapter 3. Since we wish to demonstrate that our ICA-based approach

may provide extra information for tissue classification, simulation of speckle textures with none or a little lower order statistical information (specifically FOS) would be helpful. However, complete reduction of lower order statistics by simulation is difficult. In the case of simulation of lesion region embedded in the normal tissue background, we use the preprocessed image in which some first order statistics measures (mean and variance) are intentionally extracted so that we can demonstrate that the proposed HOS method works well while the first order statistical methods are guaranteed to fail.

By simulating textures with underlying tissue properties changing gradually in a small range, we can obtain a series speckle texture images which have very similar visual patterns and statistical properties. This can be treated as a simulation of the early stage pathological deviation of tissue from normal to abnormal status with gradually changing tissue microstructures. By applying our proposed algorithm and other algorithms for comparison to this series of challenging images, we wish to demonstrate that our proposed algorithm outperform other algorithms. Each series of simulated tissues have been carefully chosen so that the *pdf* of corresponding speckle textures approach towards Rayleigh limit in a special fashion. By further extracting the first order statistics as much as possible using the previously mentioned method, we can also use these simulated textures to investigate the behavior of HOS when the underlying tissue properties change gradually.

1.2 B-mode Ultrasound Imaging

Ultrasound imaging has been successfully used as one of the primary medical imaging modalities for more than two decades. It has many advantages over other imaging modalities because it is noninvasive, non-ionizing, real-time, portable and low cost. In ultrasound imaging high frequency sound waves are generated and made to interact with human tissue. The echoed signal carrying diagnostic information is received, processed and visualized. B-mode ultrasound and Doppler ultrasound are two major classes of medical ultrasound

imaging modes. Doppler ultrasound is primarily used to measure blood flow and B-mode ultrasound is usually adopted to image echogenicity of soft tissues. B-mode ultrasound is the most widely used clinical medical ultrasound scanning mode. In this dissertation, some image processing problems related to B-mode ultrasound will be investigated.

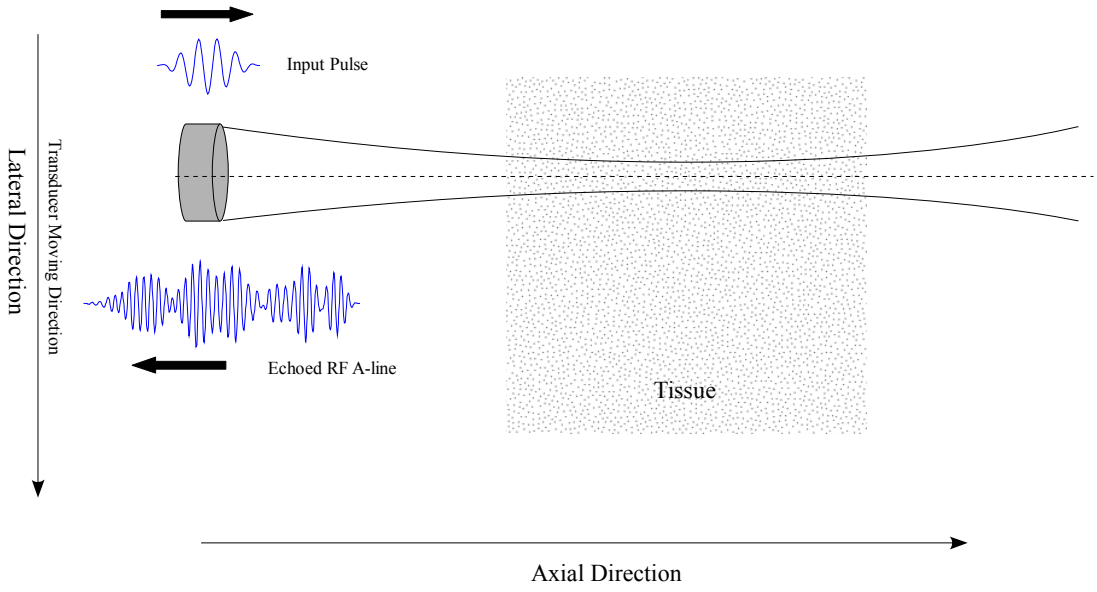


Figure 1.1: Illustration of B-mode ultrasound imaging using a single circular concave transducer. A pulse signal is sent out and an echoed RF A-line signal is received. Transducer is moved along the lateral direction to obtain multiple A-lines.

The working principle of B-mode ultrasound imaging can be illustrated in Fig. 1.1. In this simple setting, an ultra high frequency pulse signal transmitted from the aperture of a circular concave transducer propagates into the tissue located in the focal region, and interacts with point scatterers that constitute the tissue microstructure. The backscattered echo signal is received by the same transducer. The received signal is usually called an A-line which is a 1-D radio frequency (RF) signal modulated by the properties of tissue (the signal in blue color in Fig. 1.2). When the transducer is moved along the lateral direction, multiple A-lines can be obtained one at a time and then assembled later as a 2-D signal array. In practice, only the envelopes of the A-lines (the signal in red color in Fig.1.2) are stored and logarithmically transformed. Now the 2-D array consisting of multiple enve-

lope detected A-lines forms a B-mode ultrasound image (see Fig. 1.3). In real medical applications, many advanced transducers such as linear and phased array transducers may be used with various scanning geometries. The transmitting and receiving transducers can also be separately placed at different locations. B-mode ultrasound imaging system can be modeled as a linear system characterized by a corresponding point spread function (PSF) [68]. If tissue is modeled as a collection of rigid point scatterers with varying scattering strengthes (reflectivities), a RF A-line is the convolution of the PSF with the tissue scatterers. Various factors, such as the excitation pulse, the system function of transducer and the two-way diffraction effect control the shape of the PSF. Furthermore, PSF is spatially variant. It is dependent on the axial depth, i.e., distance from the transducer surface to a point in space. As a result, the ultrasound images are non-stationary along the axial direction (beam direction). Fig. 1.4 gives an example of PSF of a transducer that varies with axial depth. In reality, the PSF is a 3-D function due to the circular symmetry of the beam profile. Associated with a given 3-D PSF, the concept of "resolution volume" (or "cell volume") will be defined in the next section.

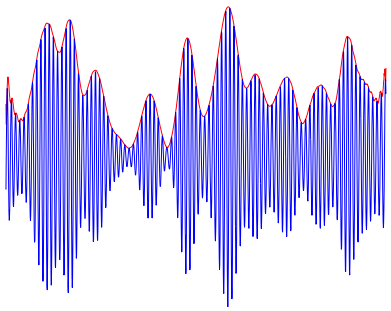


Figure 1.2: A RF A-line signal and its corresponding envelope signal shown in red color.

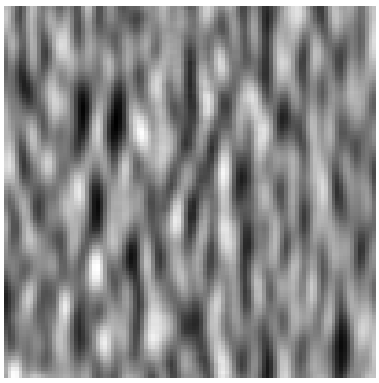


Figure 1.3: A typical ultrasound B-mode image. Each row of this image is an envelope A-line signal similar to the one shown in Fig.1.2

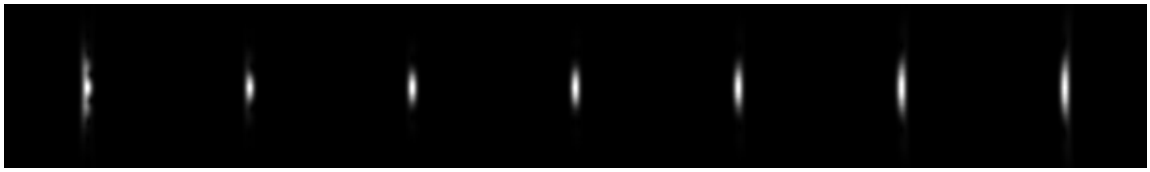


Figure 1.4: Envelope detected PSFs of a given transducer obtained at different axial depths ranging from 4cm to 10 cm.

1.3 Formation of Speckle

Due to finite aperture size (beam spreading or diffraction-limited) and finite bandwidth (short time duration pulse or band-limited) of a transducer, the ultrasound imaging system has limited resolution in both the lateral and axial direction of sound beam. Lateral resolution is affected by the diffraction-limited effect while axial resolution is controlled by the band-limited nature of the pulse. The so-called *resolution volume* (or *cell volume*) can then be approximately defined as the 3-D volume constrained by the 3dB beam profile and pulse profile. The transducer is considered a phase sensitive detector. When the echoed signals of all the random scatterers within the resolution volume are added constructively and destructively at the surface of transducer, a random interference pattern is formed in the B-mode ultrasound image that is referred as speckle [4] (see Fig.1.3).

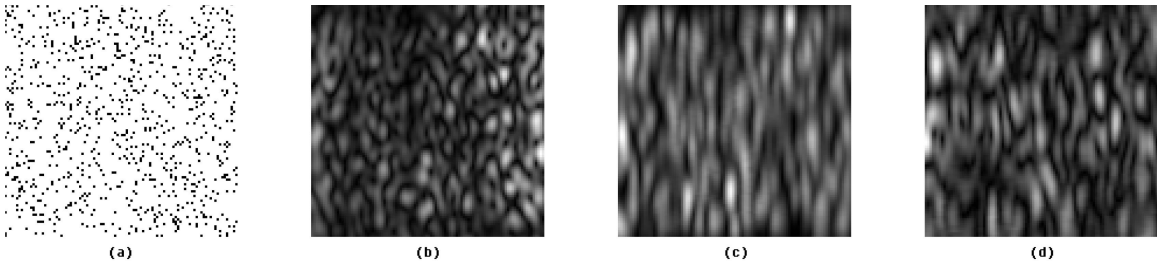


Figure 1.5: (a) Simulated tissue. (b),(c),(d) B-mode images at different axial depths (4cm, 7cm and 19cm). This figure illustrates how the speckle texture patterns, with the same underlying tissue microstructure, change with different PSFs.

The presence of speckle is the inherent nature of many coherent imaging modalities, such as ultrasound imaging and synthetic aperture radar (SAR) imaging. The statistics of speckle

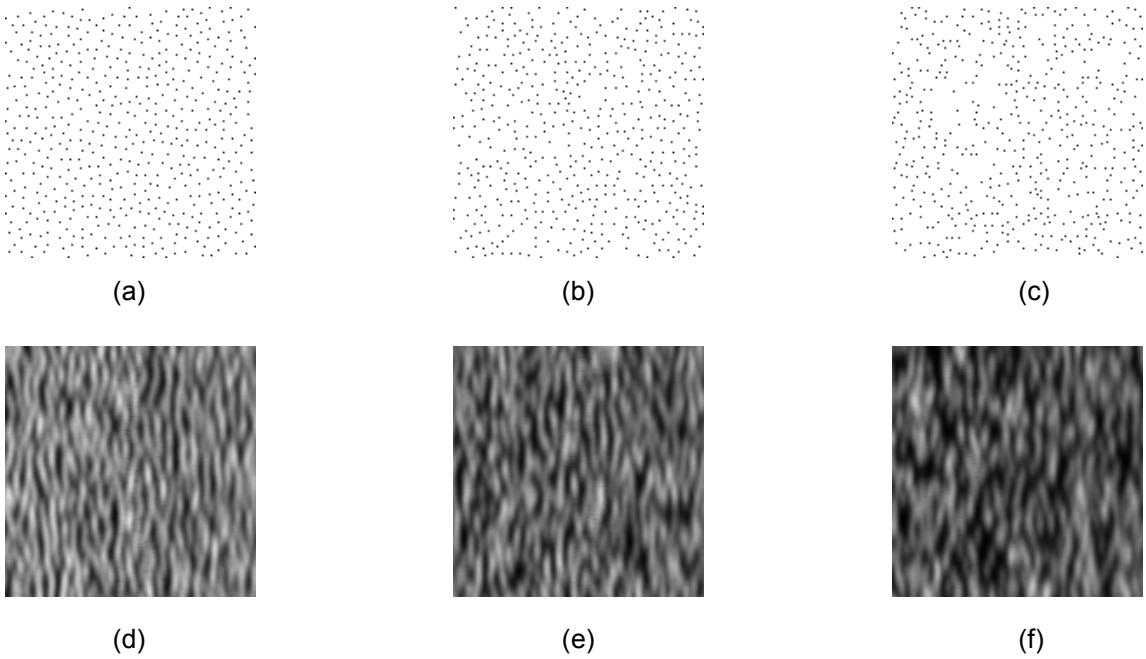


Figure 1.6: (a),(b),(c) Simulated tissues with different scatterer spatial randomness. (d),(e),(f) The corresponding B-mode images with fixed PSFs. This figure illustrates how the speckle patterns, with the same PSF, change with different underlying tissue properties. One may notice that the change of texture pattern is very slight.

depends on both the imaging system and tissue microstructure (see Fig. 1.5 and Fig. 1.6). One should note that there is no one to one spatial correspondence between the speckle texture pattern and the underlying tissue microstructure. Speckle gives ultrasound images their characteristic granular appearance so that it can be undesirable, especially when we are perceiving low contrast lesions. Hence speckle is often treated as annoying noise that needs to be reduced. On the other hand the speckle signal also carries useful information about the tissue microstructure. Both speckle texture analysis and speckle reduction, like two sides of the same coin, are important in diagnostic ultrasound. In this dissertation we will explore the application of Independent Component Analysis (ICA) to both problems.

1.4 Tissue Scattering Model

To further understand the properties of ultrasound images, it is necessary to model the scattering of tissue. Scattering is a general physical process where some forms of radiation, such as light and sound, are forced to deviate from a straight trajectory by one or more localized non-uniformities in the medium through which they pass. In ultrasound imaging, scattering is related to small inhomogeneities in acoustic impedance, which are randomly distributed in 3-D space. Usually groups of similar cells (the basic building blocks) are organized into tissues. Different type of tissue are combined to perform specific functions as an organ. According to the structure of tissues, the inhomogeneities often appear at the boundary of some small structural elements such as cells or vessels so that it is natural to use a discrete model which describes the tissue as an aggregate of small subwavelength (comparing to the wavelength of ultrasound) point scatterers. A classification scheme for tissue scattering is given in [78]. The first class of scattering is associated with relaxation phenomena of translational and rotational vibration modes of biological macromolecules. This type of scattering produces absorption and sound speed dispersion. The second class of scattering is caused by the concentrations of living cells with spacings much smaller than the resolution volume. This type of scattering would result in a component of speckle denoted as diffusive scattering. The third class of scattering is due to the structure of tissue in concentrations of a few scatterers per resolution volume. The last class of scattering is specular which is associated with organ and vessel boundaries. In this proposal, our main interest will be the speckle texture pattern caused by the second and third classes of scattering. The last class of scattering will also be considered for despeckling problem.

1.5 Ultrasound Image Processing and Tissue Characterization

Dealing with speckle is one of the major tasks of ultrasound image processing besides other tasks such as estimation of volumes and areas, measuring shapes and classifying objects. Ultrasound image processing is closely related to ultrasound tissue characterization (UTC) which can be defined as "the assessment by ultrasound of quantitative information about the characteristics of biological tissues, and pathological changes thereof" [79]. The quantitative information can be extracted from the echographic data either in the form of RF signals or B-mode envelope images. Analysis of RF signals can be used to estimate some acoustic tissue characteristics such as frequency dependence of attenuation and tissue strain. Analysis of the 2-D textures in the B-mode images provides us some information regarding the histological characteristics of tissue such as the effective scatterer number density and the spatial regularity in the histology, e.g., the lobular structure of the liver parenchyma [85]. In this proposal we will only consider the envelope detected B-mode images instead of the RF signals. The texture analysis results will be helpful for early soft tissue disease detection or ultrasound image guided biopsy.

1.6 First, Second and Higher Order Statistics

Several important concepts, in the context of image processing, need to be introduced here since they will be encountered in the later sections or chapters. For further information, readers can refer to some general statistics or image processing books such as [65].

1.6.1 First Order Statistics (FOS)

First order statistics (FOS) only consider the overall property of individual image pixels, such as mean and variance, and do not consider the relationship of neighboring pixels.

Any image processing methods based on the 1-D histogram (or *pdf*) and any of its derived measures are first order statistics based methods. Usually, an image can be modeled as a random process in which each pixel can be regarded as a random variable. Only under the assumption that all these random variables are statistically independent, first order statistics can then be used to fully characterize the image. For a real world image, the pixels usually have sophisticated dependencies with each other so the first order statistics alone may not be enough. For example, if we randomly shuffle the pixels in an image, the shuffled image still has exactly the same histogram as original image but two images have totally different appearances. For texture analysis task, the pixel dependency gives us the unique visual appearances of different textures. In chapter 2, a strict mathematical treatment of the concept "dependency" will be given by defining two concepts *uncorrelatedness* and *independence*.

1.6.2 Second Order Statistics (SOS)

Second order statistics (SOS) consider the relationship between groups of two (usually neighboring) pixels in the original image. Autocorrelation function and power spectrum are typical examples of second order statistics. For a discrete signal f[n], autocorrelation function can be defined as

$$R[k] = \mathrm{E}\{f[n]f[n+k]\} \tag{1.1}$$

where k is the time lag. When we consider a multidimensional random vector, autocorrelation matrix will be used to replace the 1-D autocorrelation function.

1.6.3 Higher Order Statistics (HOS)

In contrast with lower order statistics (FOS and SOS), higher order statistics (HOS) consider the relationship among three or more pixels. It is not commonly implemented due to calculation time and interpretation difficulty. HOS measures are extensions of second-order measures. For example, autocorrelation function defined in above can be extended to

third-order measure called third-order moment as

$$M[m,k] = \mathrm{E}\{f[n]f[n+m]x[n+k]\} \tag{1.2}$$

where m and k are two independent time lags. Some of the key advantages to higher orders statistics over traditional second order statistics are: (1) Second order measures (such as the power spectrum and autocorrelation functions) contain no phase information; (2) Since any Gaussian signal is completely characterized by its mean and variance, the HOS of the Gaussian signal are either zero (e.g. the 3rd order moment of a Gaussian signal is zero), or contain redundant information (e.g. the fourth order moment). For non-Gaussian signals, HOS may provide more information beyond the first and second order statistics. In chapter 2 we will introduce Independent Component Analysis (ICA) and explain how it is related to HOS. We will also show how ICA can be used to overcome the severe computation difficulty when dealing with HOS.

1.7 Literature Review

This literature review is divided into two parts: the first part relates to speckle texture analysis and tissue characterization, and the second part pertains to despeckling. In fact, texture analysis and despeckling are two closely related subjects in some aspects, for example, many of the despeckling techniques are based on the results of statistical analysis of speckle texture. The distinction lies in how the speckle is treated.

1.7.1 Speckle Texture Analysis and Tissue Characterization Related

A lot of methods have been proposed for ultrasound speckle texture analysis and tissue characterization in the last two decades and they can be roughly classified as following three major categories:

- *First order statistics based methods*. Several 1-D probability distribution functions have been proposed in the literature to describe the first order statistics of speckle amplitude. When the number of scatterers within the resolution volume is large enough (> 10), the amplitude of speckle obeys Rayleigh distribution [4] [86]. Under this condition, the speckle is called fully developed. Deviations from such special scattering conditions have previously been modeled via Rice distribution [33] and K-distribution [35]. Rice distribution accounts for the coherent component due to the presence of regular structure of scatterers within the tissue and K-distribution accounts for the low effective scatterer density. General models, such as generalized K-distribution [36], have also been proposed in the literature to account for a broader range of scattering conditions in a unified manner. Statistics of the logarithmically transformed B-mode images have also been investigated [16]. Many parameters can be derived from the aforementioned distributions to characterize the properties of tissue. A number of UTC algorithms based on these models have been proposed in the literature (e.g., [33][71][72]). Markov Random Fields (MRF) algorithms based on these models were also suggested for the ultrasound image texture analysis [2].

  First order statistics works well in situation when two different speckle textures have distinct first order statistical measures. In [79], the authors gave an example of using adaptive L_2-mean filter to enhance the detectability of a round lesion region embedded in the background. The lesion contrast with respect to background is +3dB. The results showed that after filtering the lesion detectability improved about 30%. However, such kind of first order statistics based algorithms are easily influenced by the factors such as contrast change and attenuation in the images. As a matter of fact, it is possible (for example, by changing the imaging condition) to find two ultrasound speckle texture images with very similar lower order moments (up to second order) but relatively distinct 2-D visual patterns. The disadvantage of a first order statistics

based algorithm is due to the fact that it completely ignores the relationship among pixels so the information containing in the textural pattern is not utilized at all.

On the other hand, some 1-D probability distributions for UTC derived from rigorous ultrasound physics models are sometimes analytically intractable. For example, the generalized K-distribution has no closed form expression for its *pdf* so it can not be used directly in algorithms based on *pdf* parametric estimation.

- *Second order statistics based methods*. Among many second order statistics based techniques, Autocovariance Function (ACVF) and Gray-Level Co-occurrence Matrix (GLCM) [23] are two representative methods used for ultrasound texture analysis. ACVF is introduced in [86] and [63] which can be considered as a measure of the average size of the speckles in B-mode images. The full-width-at-half-maximum (FWHM) of ACVF along both the axial and lateral directions have been used as texture parameters in case of fully developed speckle at focal zone [63]. An obvious drawback of using FWHM as texture parameters is that only the average speckle size along two directions (axial and lateral) are considered but the variation of speckle shape is ignored. GLCM uses a 2-D histogram to approximate the probability of occurrence of all combinations of two grey levels for a set of two pixels at a certain spatial distance. Many parameters can be derived from this matrix (for example, contrast, correlation, entropy and angular second moment) and used in UTC [88][82][73]. Other second order statistics based methods [79] applied to UTC are more or less the modified variants of the two methods mentioned above.

  Second order statistics only consider the relationship between two neighboring pixels. Existing methods, such as ACVF, work well only on the scale of speckle size. On a larger scale, the macroscopic changes due to the structural scatterers in tissue make the texture pattern more complicated. Many second order statistics such as

GLCM assume that the Julesz conjecture [41] based on human visual perception is correct. Julesz conjecture says that two textures are not preattentively discriminable if their second order statistics are identical. This theory established an important idea that texture might be modeled using lower order (first or second) statistics. But since counter-examples were found, the conjecture had been disproved [42][20]. As a result, first and second order statistics may not be sufficient to effectively discriminate two textures.

- *Multichannel filtering methods*. Multichannel filtering is another approach that has been used in ultrasound texture analysis. The basic idea of multichannel filtering is applying a set of filters (a filter bank) to the image and representing the textures by the marginal statistical measures of the filtered "sub-band" images . Wavelet transform and Gabor transform are two representatives in this line of techniques. For example, M-band wavelet transform was used in [50] for the classification of ultrasonic liver tissues. Wavelet image extension was used in [61] for the analysis and classification of infarcted myocardial tissue. Gabor transform was used in [75] for the diagnosis of prostate cancer. The drawback of most multichannel filtering techniques is that the adopted basis functions are usually analytically pre-defined and fixed. As a result, the spatial organization of speckle texture is not fully utilized to optimize the performance. Moreover, the orthogonal 2-D discrete wavelet transform (DWT) is not translation invariant as well as rotation invariant, which is not suitable for feature detection. On the other hand, although the continuous 2-D wavelet transform is translation/rotation invariant, the computational complexity is too high for practical applications. The 2-D Gabor transform encounters the similar problems. Generally speaking, how to properly choose a subset of the redundant Gabor or wavelet bases to efficiently represent the statistics of the texture remains a difficult problem. Some thorough comparison of ICA with Gabor and wavelet transform will be given in chapter 4.

In the review above, [4], [86],[33], [35], [86], [71] and [75] are identified as the key references.

1.7.2 Despeckling Related

Various linear and nonlinear filtering algorithms were developed in the past few decades for the ultrasound despeckling task. These techniques can be classified into following categories:

- *Local statistics filtering*. The working principle of this category of filters may be described by a weighted average calculation using subregion statistics to estimate statistical measures over different pixel windows. All of them assume that the speckle noise model is multiplicative in nature. Typical such filters include L2-mean filter [45], adaptive mean filter [84] and Wiener filter [18], all making use of the first order statistics such as mean and variance.

- *Geometric filtering*. The concept of the geometric filtering is that speckle appears in the image as narrow walls and valleys. The geometric filter, through iterative repetition, gradually tears down the narrow walls and fills up the narrow valleys, thus smearing the weak edges that need to be preserved [5].

- *Homomorphic filtering*. This type of filter performs image enhancement by calculating the FFT of the logarithmic compressed image, applying a denoising homomorphic filter function in frequency domain, then performing the inverse FFT of the image [69].

- *Anisotropic diffusion filtering*. Anisotropic diffusion filtering removes noise from an image by modifying the image via solving a partial differential equation (PDE). This nonlinear technique smoothes homogeneous image regions and retains image edges. It performs smoothing parallel to the edge and stops smoothing by impeding diffusion

at image edges. Applications of this technique on ultrasound despeckling can be found in [47] [39]. Properly choosing a diffusion coefficient function which can be adaptive to the edges with different contrast is the main challenge when anisotropic diffusion filtering is adopted for despeckling.

- *Wavelet filtering*. Wavelet filtering has been widely used in ultrasound despeckling [22][57]. Most algorithms are based on the soft-thresholding denoising method proposed by Donoho [14]. Since the 2-D DWT bases are pre-defined and not rotation-invariant, the statistics of the original speckle image is not fully utilized. Thus the pre-defined wavelet dictionary might not be tuned to best differentiate between signal and noise. As a result, the performance of the soft thresholding algorithm is not optimal for images [53].

The above reviewed techniques either explicitly use the first or second order statistics of the image, or implicitly use the higher order statistics (e.g., through multichannel filtering) by taking into consideration the shape of speckle and edges, but most of these techniques are marginally data adaptive. The performance of those techniques reported in different studies varies and usually are dependent on a specific clinical application. Different performance evaluation criteria are used in different studies. Many of the existing despeckling filters are sensitive to the shape and size of the window. Some of them do not enhance the edges, but only inhibit smoothing near the edges. So the speckle in the neighborhood of an edge will remain after filtering.

In the review above, [45], [84], [18], [22] and [57] are identified as the key references.

1.8 Organization of the Dissertation

The dissertation is organized as follows:

- Chapter 1 (this chapter) gives a brief introduction of ultrasound B-mode imaging and

some basic concepts such as speckle. Literatures for ultrasound texture analysis and despeckling are reviewed. This chapter also introduces the motivation of this study.

- Chapter 2 gives a brief introduction of ICA theory and its application in image processing.

- Chapter 3 introduces the numerical simulation of tissue scatterer fields and corresponding B-mode ultrasound images. Some classical 1-D probability distributions for UTC are introduced.

- Chapter 4 proposes an ICA-based multichannel filtering algorithm for ultrasound speckle texture analysis and tissue classification. The algorithm is applied to the numerically simulated B-mode speckle images and the results are compared with those obtained by using several classical methods. This chapter also investigates the performance change of higher order statistics based method when the underlying tissue properties change in an gradual fashion.

- Chapter 5 proposes a despeckling algorithm by adapting the ICA sparse code shrinkage (ICS-SCS) with a preprocessing step considering the speckle properties of ultrasound images. Experimental results are presented and compared with those obtained by using several classical methods.

- Chapter 6 summarizes the contributions of this work, draws some conclusions and gives recommendations for future research.

1.9 Publications Based on the Dissertation

The following papers have been published based on this dissertation:

1. Di Lai, Navalgund Rao, Chung-hui Kuo, Shweta Bhatt, and Vikram Dogra. "Independent Component Analysis Applied to Ultrasound Speckle Texture Analysis and Tissue

Characterization", *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, Cité Internationale, Lyon, France, August 23-26, 2007

2. Di Lai, Navalgund Rao, Chung-hui Kuo, Shweta Bhatt, and Vikram Dogra, "An Ultrasound Image Despeckling Method Using Independent Component Analysis", *Proceedings of 2009 IEEE International Symposium on Biomedical Imaging (ISBI 09)*, Boston, MA, June 28 - July 1, 2009

Chapter 2

Independent Component Analysis (ICA) and Its Application

2.1 Introduction of ICA

Independent Component Analysis (ICA) is a relatively new statistical computing tool for revealing the hidden factors that underlie sets of random variables, measurements or signals. It defines a generative model for the observed multivariate data which are assumed to be linear or nonlinear mixtures of some unknown latent variables, while the mixing system itself is unknown. The latent variables are assumed to be non-Gaussian and mutually independent, and they are called the independent components (ICs) of the observed data. ICA has been successfully applied to the real world problems such as biomedical signal processing, audio signal separation, telecommunication, etc. ICA is closely related to Blind Source Separation (BSS) which can be formulated as the problem of separating or estimating the waveforms of the original sources from an array of sensors or transducers without knowing the characteristics of the transmission channels. While ICA methods assume the original sources are as independent as possible and use higher order statistics, BSS methods generally do not require the original sources to be completely independent and may use only second order statistics [11]. In this chapter, we will give a brief introduction of ICA

and its application on image processing.

2.1.1 Independence, Uncorrelatedness and Whiteness

Before introducing the basic ICA model, we need to recall some basic definitions. Let's consider some random variables x_i ($i = 1, \cdots, M$) with joint density $p(x_1, x_2, \cdots, x_M)$. For simplicity, assume that each of those variables has zero mean. These variables are said to be mutually *independent* (or more strictly, *statistically independent*), if their joint density function can be factorized as:

$$p(x_1, x_2, \cdots, x_N) = p_1(x_1)p_2(x_2)\cdots p_M(x_M) \tag{2.1}$$

where $p_i(x_i)$ denotes the marginal density of x_i. Independence must be distinguished from *uncorrelatedness*. Two random variables are said to be *uncorrelated* if

$$\mathrm{E}\{x_m x_n\} = \mathrm{E}\{x_m\}\mathrm{E}\{x_n\}, \qquad \text{for} \quad m \neq n \tag{2.2}$$

Independence (or strictly, statistical independence) is in general a much stronger condition than uncorrelatedness. Independent variables are always uncorrelated but not vice versa. Only for multivariate Gaussian variables, uncorrelatedness is equivalent to independence. Indeed, if x_i ($i = 1, \cdots, M$) are independent then

$$\mathrm{E}\{g_1(x_m)g_2(x_n)\} = \mathrm{E}\{g_1(x_m)\}\mathrm{E}\{g_2(x_n)\}, \qquad \text{for} \quad m \neq n \tag{2.3}$$

where g_1 and g_2 can be any absolutely integrable nonlinear functions. Uncorrelatedness can be defined as a special case when g_1 and g_2 are linear functions in the above equation.

Whiteness is another important concept closely related to uncorrelatedness. A zero-

mean random vector $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_M)^T$ is said to be *white* if

$$\mathrm{E}\{\mathbf{x}\mathbf{x}^{\mathrm{T}}\} = \mathbf{I} \tag{2.4}$$

where $\mathbf{I}$ is an identity matrix. A random vector can be whitened by applying a proper linear transform (decorrelating) followed by a scaling or sphering step (variance normalization).

2.1.2 Basic Linear ICA Model

Assume an observed random vector $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_M)^T$ (each element of the $\mathbf{x}$ is a scalar random variable) can be modeled as the linear combination of some latent variables s_j:

$$x_i = \sum_{j=1}^{N} a_{ij} s_j, \qquad i = 1, 2, \ldots M \tag{2.5}$$

or expressed in matrix form

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{2.6}$$

where the unknown mixing matrix $\mathbf{A}$ is a constant matrix with M rows and N columns. We assume $\mathbf{x}$ and $\mathbf{s}$ are centered (mean subtracted). Here s_j are latent random variables called independent components (or in abbreviated form, ICs). For simplicity, here we just assume $M = N$ in this basic model (other cases will be discussed later). When $M = N$, the unknown mixing matrix $\mathbf{A}$ is square. In other words, the number of independent components is equal to the numbers of observed mixtures. If we also assume that $\mathbf{A}$ is invertible, then we can compute its inverse to obtain independent components ($\mathbf{s} = \mathbf{A}^{-1}\mathbf{x}$). If $\mathbf{A}$ is not invertible, that means there are redundant mixtures that could be omitted, in which case the matrix would not be square. Then we just consider the cases when $M \neq N$.

ICA aims to estimate both the unknown $\mathbf{A}$ and unknown $\mathbf{s}$, observing only $\mathbf{x}$. However, some assumptions must be made here: the s_j are mutually independent and non-Gaussian (with the possible exception of only one component to be Gaussian). Most ICA algorithms

are searching for a separation matrix $\mathbf{W}$ to demix the data based on various estimation principles of independence. The demixing model can be written as

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} \tag{2.7}$$

where $\hat{\mathbf{s}}$ is the estimation of $\mathbf{s}$ and $\mathbf{W}$ is the inverse (or pseudo inverse) of $\mathbf{A}$. Usually, the independent components s_j can be estimated up to some multiplicative constant and the order is not determined.

When $M > N$, the row dimension M of $\mathbf{A}$ can be reduced to N to make $\mathbf{A}$ to be square. Principal Component Analysis (PCA) is a technique usually used for this redundancy reduction. We will discuss PCA in Section 2.1.6. $M > N$ is a usual case in practical problems since observed variables are usually the linear combination of a few latent sources.

When $M < N$, the number of mixture x_i is smaller than the number of independent components. This means that the mixing system is not invertible. This case is usually called *ICA with overcomplete bases*. ICA with overcomplete bases is much more difficult than the basic ICA problem, and the estimation methods are less developed. The simplest way of estimating the independent components is using pseudoinverse of the mixing matrix. This yields

$$\hat{\mathbf{s}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{x} \tag{2.8}$$

In some situations such a simple pseudoinverse gives a satisfactory solution, but in many cases we need a more sophisticated estimate. Several advanced algorithms (e.g., using maximum likelihood estimation) have been developed to deal with this overcomplete bases situation. Readers can refer to [30] for further information.

2.1.3 Independence and non-Gaussianity

For a random variable, non-Gaussianity can be roughly defined as the deviation of its *pdf* from Gaussian distribution. Non-Gaussianity is of paramount importance in ICA estima-

tion and can be related to independence by the central limit theorem (CLT), which is a classic result in probability theory. A short introduction of CLT is given here. Let

$$x_k = \sum_{i=1}^{k} z_i \tag{2.9}$$

be a partial sum of a sequence $\{z_i\}$ of independent and identically distributed (*i.i.d*) random variables z_i. Since the mean and variance of x_k can grow without bound as $k \to \infty$, consider instead of x_k the standardized variables

$$y_k = \frac{x_k - m_{x_k}}{\sigma_{x_k}} \tag{2.10}$$

where m_{x_k} and σ_{x_k} are the mean and variance of x_k. It can be shown that the distribution of y_k converges to a Gaussian distribution with zero mean and unit variance when $k \to \infty$. Several different forms of the theorem exist, where the *i.i.d* assumptions have been weakened. CLT can be generalized to *i.i.d* random vectors $\mathbf{z}_i$ with common mean $\mathbf{m_z}$ and covariance matrix $\mathbf{C_z}$.

Loosely speaking, the sum of two independent random variables usually has a distribution that is closer to Gaussian than any of the two original random variables [30]. In the basic ICA model 2.5, x_i, the linear combinations of a limited number of independent components s_j tend to be less independent and closer to Gaussian. To estimate independent components s_j, we are motivated to find certain linear combinations of x_i with maximized non-Gaussianity using certain measures such as kurtosis. The normalized kurtosis of a random variable x is defined as

$$Kurt(x) = \frac{\mathrm{E}\{x^4\}}{[\mathrm{E}\{x^2\}^2]} - 3 \tag{2.11}$$

The kurtosis of a Gaussian random variable is equal to 0. A random variable with kurtosis greater than 0 is called a super-Gaussian while a random variable with kurtosis less than

0 is called a sub-Gaussian. Usually the *pdf* of a super-Gaussian has a sharper peak and longer tail than a Gaussian *pdf* and a sub-Gaussian *pdf* is flatter than a Gaussian *pdf*. Here we assume all the variables have unit variance.

Besides maximizing non-Gaussianity, there are some other approaches for ICA estimation, such as minimization of mutual information or maximum likelihood estimation. For those broader topics please refer to [30].

2.1.4 ICA Algorithms

In Eq. 2.7, each independent component s_j is expressed as the linear combination of the observed random variables x_i. Estimation of the ICA model turns out to be an optimization problem looking for the a proper linear transform that would make s_j as independent as possible. An ICA algorithm can be generally formed by choosing an objective function (or cost function) and then minimizing or maximizing it. As introduced in 2.1.3, kurtosis can be used as such kind of objective function. In fact, there are also some other measures for maximizing non-Gaussianity such as negentropy [30].

One can use any of the classical methods of optimization, such as gradient descent method, Newton-like method, etc., for optimizing the chosen objective function. The basic ICA estimation procedure by maximization of non-Gaussianity is given as follows:

1. Whiten the data vector $\mathbf{x}$ to give $\mathbf{z}$.
2. Set iteration count $i = 1$.
3. Take a random vector $\mathbf{w}_i$.
4. Maximize non-Gaussianity of $\mathbf{w}_i^T\mathbf{z}$ under constraints $\|\mathbf{w}_i\|^2 = 1$ and $\mathbf{w}_i^T\mathbf{w}_j = 0$ $(j < i)$ by a suitable algorithm.
5. Increment iteration count by 1, go back to 3.

Or alternatively, maximize all the $\mathbf{w}_i$ in parallel, keeping them orthogonal.

In the pseudo code above, the "suitable algorithm" we chose to use is called FastICA, which was developed by Hyvärinen, *et al.* [28]. It is an efficient fixed-point iterative algorithm using kurtosis for non-Gaussianity maximization. This algorithm has been widely used in the ICA community due to its fast convergence rate and reliability. Please refer to Appendix C for details.

2.1.5 ICA and PCA

ICA is closely related to Principal Component Analysis (PCA) [40]. Actually ICA could be regarded as the generalization of PCA, which is based on second order statistics to find first N components with largest variances (principal components or PCs) by decorrelation. The purpose of PCA is to find a smaller set of variables with less redundancy that would give a representation of original multivariate data as good as possible. For PCA the redundancy is measured by correlations between data elements, while for ICA the redundancy is measured by independence. For Gaussian sources, uncorrelatedness is equivalent to independence but for non-Gaussian sources uncorrelatedness is a weaker condition than independence. Fig. 2.1 illustrates the results of applying both ICA and PCA to three linearly mixed images. It shows that ICA can recover the original images better than PCA because all the original images are non-Gaussian sources.

Let's examine a simple PCA algorithm which uses the eigenvalue decomposition (EVD) [30] of the covariance matrix. Assume we have a random vector $\mathbf{x}$ with N elements and the vector has T samples $\mathbf{x}(1), \cdots, \mathbf{x}(T)$. The elements are mutually correlated. Given sufficiently large samples of vector $\mathbf{x}$, the covariance matrix can be estimated from the samples and expressed as

$$\mathbf{C} = \mathrm{E}\{\mathbf{x}\mathbf{x}^T\} = \mathbf{E}\mathbf{D}^T\mathbf{E} \tag{2.12}$$

where $\mathbf{E}$ is the orthogonal matrix of eigenvectors of $\mathbf{C}$ and $\mathbf{D}$ is the diagonal matrix of its eigenvalues ($\mathbf{D} = \mathrm{diag}(d_1, \cdots, d_N)$). $\mathbf{D}$ and $\mathbf{E}$ can be computed by solving the eigenvector-

Figure 2.1: The demixing results of applying ICA and PCA to three linearly mixed images. The demixed images have been properly rescaled for display purposes. From left to right, Column 1: three original images, Column 2: three mixed images, Column 3: demixed images using ICA, Column 4: demixed images using PCA. The order of the demixed images have been adjusted to align with the original images for comparison purposes. This example was generated by the author.

eigenvalue problem for $\mathbf{C}$ (which is a symmetric matrix) using proper algorithms introduced in [66] or other standard numerical analysis packages. Then we sort the diagonal values (eigenvalues) in $\mathbf{D}$ in descent order and only keep the first K largest values so as to reduce dimension. The corresponding eigenvectors can be selected from $\mathbf{E}$. Now we can compute the whitening matrix $\mathbf{V}$ as

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{E}^T \tag{2.13}$$

where $\mathbf{D}^{-1/2}$ is computed by a simple componentwise operation as

$$\mathbf{D}^{-1/2} = \text{diag}(d_1^{-1/2}, \ldots, d_K^{-1/2}) \tag{2.14}$$

where $d_1, \cdots, d_K$ are eigenvalues. Then the principal components of $\mathbf{x}$ can be obtained by

$$\mathbf{z} = \mathbf{V}\mathbf{x} \tag{2.15}$$

Here $\mathbf{z}$ has been whitened (decorrelated and variance normalized) and dimension reduced. This algorithm is the closed form realization of PCA. There are many other ways to realize PCA due to different practical considerations such as the computation load, speed and numerical stability. Usually, the first step to do ICA is to take the mean away and normalize the variance to unit. Then PCA can be used as the whitening step to make the mixing matrix to be square, reduce noise and prevent over-learning.

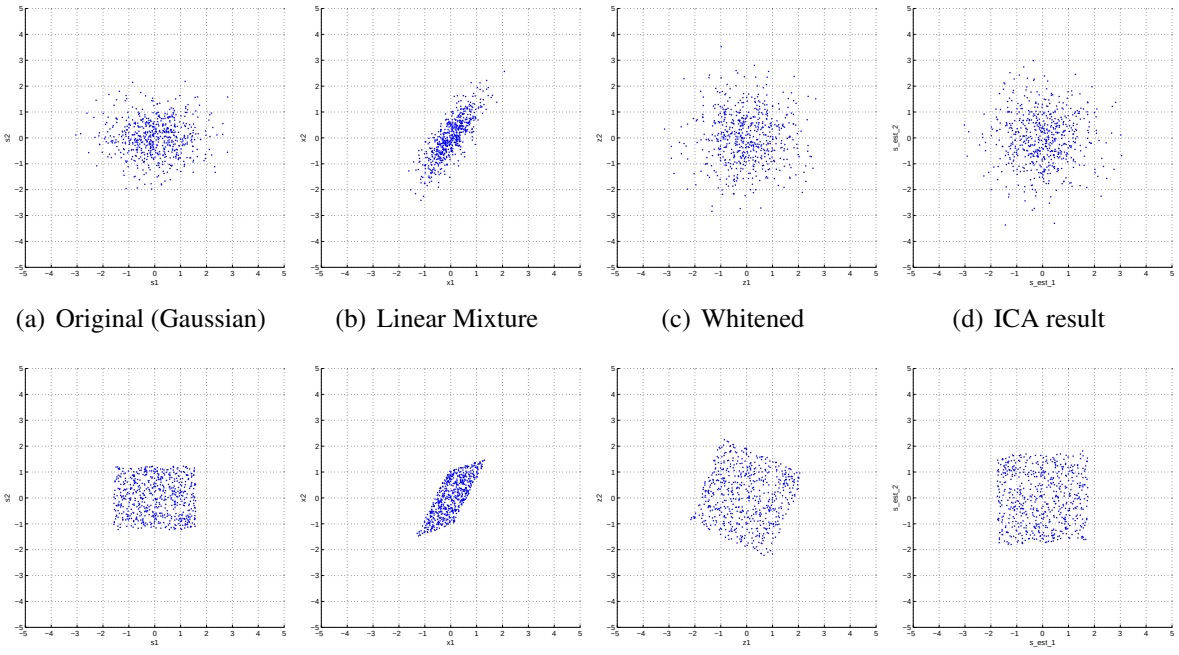
2.1.6 Why Gaussian is not allowed in ICA?

Whitening is only half of ICA. If an orthogonal transformation $\mathbf{U}$ is applied to $\mathbf{z}$,

$$\mathbf{y} = \mathbf{U}\mathbf{z} \tag{2.16}$$

one can easily prove that the transformed vector $\mathbf{y}$ is also white (see Appendix A). This means that we can not tell if the independent components are given by $\mathbf{z}$ or $\mathbf{y}$ using the whiteness property alone. Whitening helps us understand why Gaussian variables are forbidden in ICA (actually ICA only allows at most one IC to be Gaussian). Let's assume two independent Gaussian source variables s_1 and s_2 are linearly mixed. We also assume the mixing matrix $\mathbf{A}$ is orthogonal (because the data has been whitened). One can prove that the joint distribution $p(x_1, x_2)$ does not change with any orthogonal mixing matrix [30] (a proof is given in Appendix B). This indicates that we can not estimate the matrix $\mathbf{A}$.

Figure 2.2 illustrates the conclusion in above by showing the scatter graphes of each step of the demixing process using ICA. Both Gaussian mixture and non-Gaussian mixture are considered. Gaussian mixture is generated from the linear combination of two independent Gaussian random variables. Non-Gaussian mixture is generated from the linear combination of two independent non-Gaussian random variables (using uniform distribution). The mixing matrix is a randomly generated 2 by 2 constant matrix. In Gaussian mixture case (first row of Fig. 2.2), we can see that although the whitening process can decorrelate the mixture of Gaussian variables, the ICA transform is not able to rotate the coordinates to identify the independent components. So the original signals can not be recovered. In non-Gaussian case (second row of Fig. 2.2), since the source signals are non-Gaussian, the ICA transform has successfully found a proper coordinate system to approximately recover the original signals (up to some multiplicative constant).



(a) Original (Gaussian) (b) Linear Mixture (c) Whitened (d) ICA result

(e) Original(non-Gaussian) (f) Linear Mixture (g) Whitened (h) ICA result

Figure 2.2: Comparison of the demixing results of Gaussian mixture and non-Gaussian mixture using ICA. From (a) to (d): Original Gaussian variables, corresponding linear mixture, whitened variables and ICA result. From (e) to (h): Original non-Gaussian variables, corresponding linear mixture, whitened variables and ICA result. This example was generated by the author.

2.2 ICA Applied to Image Processing

ICA has been applied to the image processing tasks such as feature extraction and denoising. A fundamental approach in signal processing is to design a statistical generative model of the observed signals. The components in the generative model then give a representation of the data. Such a representation can then be used in such tasks such as compression, denoising and pattern recognition. Image representation is often based on discrete linear transformations of the observed data. Consider a black-and-white image whose gray-scale value at the pixel indexed by x and y is denoted by I(x, y). Many basic models in image processing express the images I(x,y) as a linear combination of some features or basis functions $a_i(x, y)$:

$$I(x, y) = \sum_{j=1}^{n} a_j(x, y)s_j \tag{2.17}$$

where s_j are stochastic coefficients, different for each image $I(x, y)$, and a_j are called image bases. Figure 2.3 shows an example of this model. This kind of linear combination model gives a useful description of an image on the low level. In practice, we apply Eq. 2.17 to an image window (for example, 8×8 window) to obtain the bases that give a compact expression of the image at the scale of that window size. Alternatively, we can reshape the image window into a 1-D column vector $\mathbf{x} = (x_1, x_2, \cdots, x_M)^T$ in which case we can express the representation as

$$\mathbf{x} = \sum_{j=1}^{n} \mathbf{a}_j s_j = \mathbf{As} \tag{2.18}$$

This is the basic ICA model that we defined in Eq.2.6. The data vector $\mathbf{x}$ has been expressed as the linear combination of column vectors of the mixing matrix $\mathbf{A}$. The column vectors $\mathbf{a}_j$ are usually called bases while the row vectors $\mathbf{w}_i$ of the demixing matrix $\mathbf{W}$ are usually called filters (because the observed signals are projected onto those vectors to obtain ICs). If we assume $\mathbf{a}_i$ forms an invertible linear system, that is, matrix $\mathbf{A}$ is square, then we can

easily get the relationship between $\mathbf{a}_i$ and $\mathbf{w}_i$ as:

$$\mathbf{a}_i = \mathbf{A}\mathbf{A}^T\mathbf{w}_i = \mathbf{C}\mathbf{w}_i \tag{2.19}$$

which shows that the basis vectors $\mathbf{a}_i$ are obtained by filtering the coefficients in $\mathbf{w}_i$ by the filtering matrix given by the autocorrelation matrix [30]. For natural image data, the autocorrelation matrix is typically a symmetric low-pass filtering matrix, so the basis vectors $\mathbf{a}_i$ are basically the smoothed version of the filters $\mathbf{w}_i$.

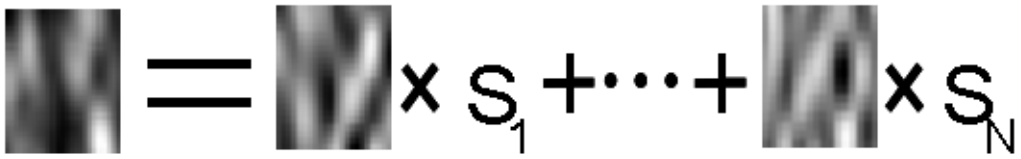


Figure 2.3: An image window expressed as the linear combination of some basis functions. If the basis functions are obtained by ICA, the coefficients s_j are independent components.

The ICA basis functions for image processing are usually learned from the training samples. This learning process actually attempts to estimate the building blocks of the images at low level. If an observed image window is projected into those bases, the image pixel values will be transformed into independent component values. The bases obtained by ICA have some special advantages over many conventional bases such as Gabor and wavelet. More discussions about this will be given in Chapter 4.

2.2.1 ICA and Sparse Coding

ICA is closely related to another technology called sparse coding. In sparse coding, a data vector is represented by a set of basis vectors so that only a small number of basis vectors are activated at the same time (i.e., sparsity). In neural network interpretation, each basis vector corresponds to one neuron and coefficients s_j are given by the activations. A given

neuron is activated only rarely. This means the distribution of s_j has a sparse distribution which usually has a probability density with a peak at zero and heavy tails. Laplacian (or double exponential) density is such an example. Sparseness is generally equivalent to the concept of super-Gaussianity. Obviously sparse coding can be fit into the framework of ICA. One could estimate a linear sparse coding transformation of the data by formulating a measure of spareness of the components, and maximizing the measure in a set of linear transformations. Maximizing sparsity is one method of maximizing non-Gaussianity (here is super-Gaussianity). Sparse coding can be applied to image processing problems such as compression and denoising. For denoising, one can use thresholding methods to set those components that contribute to noise to zero. Hyvärinen *et al.* [27] have successfully applied this method to the denoising of natural images.

2.2.2 Applications

For general image processing, ICA has been applied to the subjects such as image feature extraction [62] [10], texture analysis and classification [52] [70] [21] [54] [49], edge detection [89] [8] and blind deconvolution [43] [55] [24]. Most of those applications utilize the basis functions learned from training samples to form a generative model for the image. Then different domain-specific algorithms were used for image processing or pattern recognition based on the model.

ICA has been widely used in the field of remote sensing imaging for speckle reduction, feature extraction and data fusion [17]. Karvonen [44] has successfully applied ICA on the sea ice SAR (Synthetic Aperture Radar) image classification problem. Zhang [90] proposed a new ICA method based on high order joint cumulant applied to the image processing problem of SAR and airborne visible infrared imaging spectrometer (AVIRIS). Many ideas developed in the remote sensing field could be useful for tackling ultrasound image processing problems because remote sensing imaging also needs to deal with speckle noise, although the nature of speckle in these two fields may be different.

ICA has also found some applications in the field of medical imaging. Mayo *et al.* [56] has applied ICA to the denoising of mammographic images. Li [51] proposed an ICA method with feature selective filtering and applied it to the estimation of functional activation regions in functional MRI (fMRI). But unfortunately in the ultrasound imaging field, very few efforts have been made to explore the application of ICA. Chen's paper [7], among a few papers of ICA application on ultrasound imaging, defined an edge subspace to reduce the speckle noise in ultrasound nondestructive evaluation (NDE) C-Scan images while enhancing the edge. However, there are still not many papers dedicated to B-Scan ultrasound image texture analysis and despeckling.

Chapter 3

Numerical Simulation of B-Mode Ultrasound Images

This chapter is about the numerical simulation of B-Mode ultrasound images that will be utilized by the image processing tasks in Chapter 4 and 5. The numerical simulation consists of two parts: (1) The first part is the simulation of tissues. Based on the classical ultrasound tissue scattering model, a tissue can be simulated as a homogenous scatterer field with some parameters controlling its properties. (2) The second part is the simulation of ultrasound B-Mode imaging. The simulated tissues in (1) will be used as sources for B-Mode imaging simulation. Simulated B-Mode images are then obtained. In the last section, we give a short introduction to the classical first order statistical models for ultrasound speckle. Those models will be encountered in the later chapters and they play important roles in helping us to set the parameters of simulated tissues.

There are several reasons why we rely on the numerical simulation results instead of using real ultrasound images. First, by simulation, a large number of speckle texture images of different types can easily be obtained for the ICA training purpose. Otherwise it will become a forbidding job, as restricted by the practical conditions of this study, to collect so many clinical images or fantom images. Second, by simulation, we can easily control the underlying properties of tissues as well as the imaging conditions. So different types of

"tissues" with properties matching the prediction of classical models can be generated for research purposes. Third, the known ground truth of tissue properties makes it possible for us to cross-validate the efficiency of our proposed algorithms.

One should note that this study is a general purpose investigation of the application of ICA to ultrasound speckle textures. We do not target our work on any specific tissues such as liver tissue or kidney tissue. Real clinical applications may benefit from this kind of general purpose study but we still have a long way to go, considering many complicated practical issues. We will discuss how realistic the simulations could be in the later sections and give some comments on the possible path for leading the simulation based work to any future clinical applications.

3.1 Numerical Simulation of Tissue

A tissue scattering model has been discussed in section 1.4. For the speckle texture image analysis and tissue characterization work in Chapter 4, only the second and third classes of scattering will be considered. The second class of scattering can be simulated as scatterer with large scatterer number density (SND) so that the inter-scatterer spacing is much smaller than the resolution volume. The third class of scattering can be simulated as scatterers with smaller number density so that there are only a few scatterers per resolution volume. The resolution volume size can be roughly estimated by calculating the spatial correlation length along both the axial or lateral directions of the ultrasound beam. For the despeckling work in Chapter 5, some strong structural components are added to represent the specular scattering associated with organ and vessel boundaries.

Both the second and third classes of scattering can be simulated as a homogeneous scatterer field in the 2-D or 3-D space with certain spatial distributions. For us, only the 2-D case is considered for the sake of simplicity and because of the axial symmetry of a circular transducer. Usually there are three factors that can be used to control the tissue properties,

including the scatterer spatial distribution, scatterer number density and scatterer reflection strength (or reflectivity) distribution. Those factors are not totally unrelated to each other. For example, the situation when the scatterers become clustered or the variance of reflectivities becomes very large is equivalent to the case when the scatterer number density is very low. Any change of those factors can be regarded as the change of tissue properties so that we can use them to control the tissue simulation.

Human soft tissue has complicated microstructure. The interaction mechanism between the ultrasound wave and the tissue microstructure has always been a difficult problem which is still not fully understood. Generally speaking, the ultrasound wave reflection happens at the inhomogeneities of acoustic impedance. Those inhomogeneities are discrete in nature so it is natural to simulate the tissue as an aggregate of point scatterers. Simulating tissue as a continuous function is computationally too difficult and unnecessary. The simulated tissue scatterer field may not show the exact microstructure found in a real tissue under the microscope. However, one can tune the parameters of the simulated scatterer field so that the corresponding simulated B-Mode image shows similar statistical properties as its realistic counterpart. By analyzing the simulated B-Mode images, we may extract some features that can be correlated with those underlying parameters so we can judge the properties of the tissue. That is the goal of UTC. This argument hints that the simulated tissue scatterer field (via simulated B-Mode images) can be used as a realistic model of the real tissue.

Several approaches have been taken to simulate homogenous scatterer field with different combinations of a few controlling parameters. In our simulation, the probability distribution of scatterer reflectivity is always simulated as a Gaussian with certain mean and standard deviation. Since the changing of scatterer number density is a trivial task, in this section we will concentrate on the simulation of scatterer fields with different spatial distributions. Three methods are given as follows to simulate scatterer spatial distributions ranging from fully random ones to quasi-regular (or quasi-periodical) ones.

3.1.1 Full Spatial Randomness by Homogeneous Poisson Process

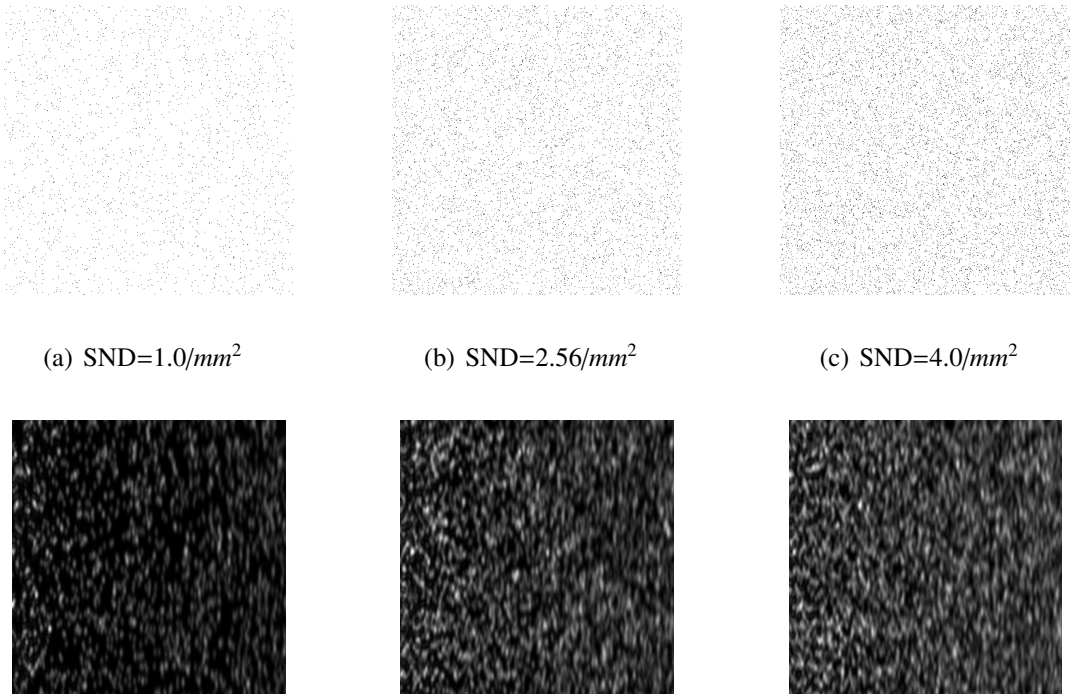
A homogeneous Poisson process [12] is characterized by a rate parameter λ, also known as intensity, such that the number of events in time interval $(t, t+\tau]$ follows a Poisson distribution with associated parameter $\lambda\tau$. This relation is given as

$$P[N(t+\tau) - N(t) = k] = \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!}, \qquad k = 0, 1, 2, \ldots, \tag{3.1}$$

where $N(t+\tau) - N(t)$ describes the number of events in time interval $(t, t+\tau]$, k is number of events, and λ is the expected number of events that occur per unit time. For the 2-D random field, the time interval is replaced by the area of spatial region and the number of events becomes the number of scatterer in the region. 2-D homogeneous Poisson process can be used to simulate scatterer field with complete spatial randomness. The points occur uniformly (i.e., homogeneously) throughout the region. There is a constant expected number density for a given unit spatial region. 2-D homogeneous Poisson process can be simulated by placing the scatterers randomly on spatial locations that obey 2-D uniform distribution. Figure 3.1 shows three simulated tissues generated by homogenous Poisson process and their corresponding B-scan images.

3.1.2 Isotropic Spatial Quasi-Regularity by Poisson Disk Process

When a constraint of minimum nearest neighbor distance d_{min} is added to the homogeneous Poisson process we obtain the Poisson disk process [12], which can be used to model a spatially quasi-regular scatterer field. The minimum distance d_{min} controls the randomness. Large d_{min} makes the random field very regular while small d_{min} makes the field more random. When $d_{min} \to 0$, the Poisson disk process becomes the homogeneous Poisson process. A scatterer field generated by Poisson disk process is spatially isotropic and shows the nature of blue noise in its frequency spectrum. Half of d_{min} is called the radius r of Poisson disk process.

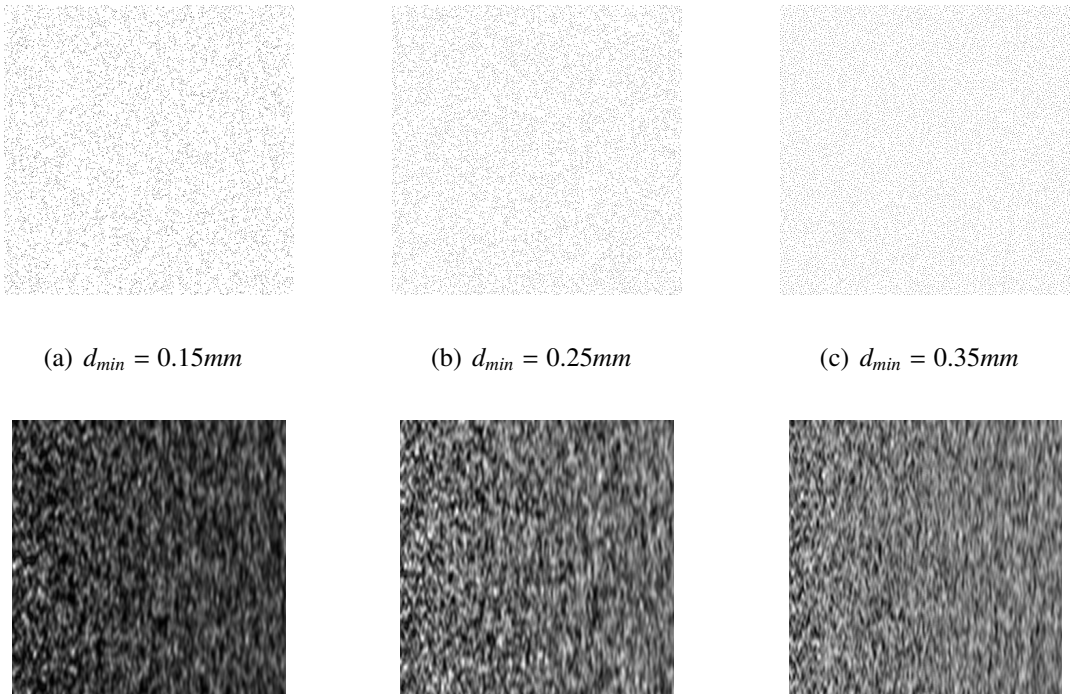


(a) SND=1.0/mm^2 (b) SND=2.56/mm^2 (c) SND=4.0/mm^2

(d) B-scan image of (a) (e) B-scan image of (b) (f) B-scan image of (c)

Figure 3.1: Three simulated tissues using homogenous Poisson process, with different scatterer number densities, and their corresponding B-scan images. The size of tissue is $50mm \times 50mm$.

The Poisson disk process was introduced in the field of computer graphics to solve the aliasing problem and has been identified as one of the best sampling patterns. There are many methods for generating Poisson disk random fields [48]. We just simply take the "dart throwing" approach by randomly (via 2-D uniform distribution) placing a point in the 2-D region and deciding whether to accept or reject it by calculating its nearest neighbor distances with the points that have already been generated. This process continues until no more points can be added. The nearest neighbor distance d_{min} can not be too large or else it will be difficult for the program to converge. The Poisson disk process is a simple way to model the spatial correlation of scatterers in the first order sense. Larger d_{min} indicates the self-avoiding nature of the scatterers, which is a phenomenon frequently encountered in the biological world. For example, the retina of the eye are distributed according to a Poisson



(a) $d_{min} = 0.15mm$ (b) $d_{min} = 0.25mm$ (c) $d_{min} = 0.35mm$

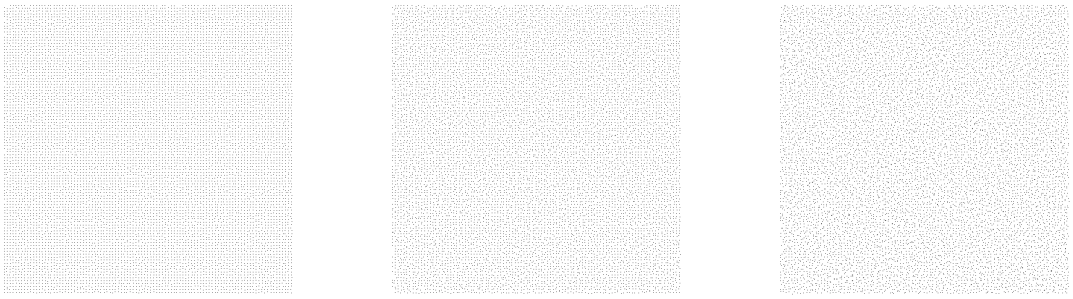
(d) B-scan image of (a) (e) B-scan image of (b) (f) B-scan image of (c)

Figure 3.2: Three simulated tissues using the Poisson disk process, with different minimum inter-scatterer distance d_{min}, and their corresponding B-scan images. The scatterer number density is fixed to $4.0/mm^2$. The size of tissue is $50mm \times 50mm$.

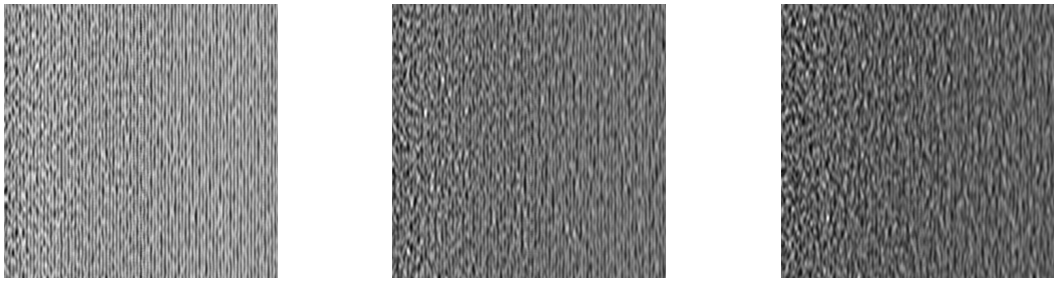
disk process [31, 32]. Fig. 3.2 shows some simulated tissues generated by the Poisson disk process and their corresponding B-scan images.

3.1.3 Anisotropic Spatial Quasi-Regularity by Random Perturbation

The random scatterer field generated by the Poisson disk process above is spatially isotropic. In order to simulate certain tissue microstructure with some quasi-periodic nature along a particular direction, we place the points on the anisotropically regular (such as rectangular) lattice, then add a random perturbation to each point. Different types of regular grid can be used. The angle θ and length d of the perturbation vector control the randomness of the point field. We assume that both θ and d are uniformly distributed in some range. Figure 3.3 shows three simulated tissues generated by random perturbation added on a rectangular



(a) $\bar{d} = 0.1mm$ (b) $\bar{d} = 0.2mm$ (c) $\bar{d} = 0.3mm$



(d) B-scan image of (a) (e) B-scan image of (b) (f) B-scan image of (c)

Figure 3.3: Three simulated tissues by using random perturbation approach, with different mean value $\bar{d}$ of random perturbation vector length, and their corresponding B-scan images. The angle of perturbation vectors is uniformly distributed in $[0, 2\pi]$. The scatterer number density is fixed to $4.0/mm^2$. The size of tissue is $50mm \times 50mm$.

lattice, and their corresponding B-scan images.

3.2 Numerical Simulation of B-Mode Imaging

For B-mode imaging simulation, Field II software developed by J. A. Jensen [37] was used. Field II has been widely used in the ultrasound imaging research community (e.g., [60]) and can yield accurate results when compared to real measurements [78]. Field II uses the concept of spatial impulse response, which is based on the Tupholme and Stepanishen approach [80, 76, 77], assuming linear propagation, to calculate the emitted ultrasound field.

Ultrasound imaging system is usually treated as a linear system [38] which is fully

characterized by its impulse response. The spatial impulse response gives the emitted ultrasound field at a specific point in space as a function of time, when the transducer is excited by a Dirac delta function. The ultrasound field for any kind of excitation pulse can then be found by just convolving the spatial impulse response with the excitation function. The impulse response varies as a function of position relative to the transducer. The total received response in pulse echo can be found by convolving the transducer excitation function with the transfer function of the transmitting transducer, then with the spatial impulse response of the transmitting aperture, then with the spatial impulse response of the receiving aperture, and finally with the transfer function of the receiving transducer. If a single transducer is used for both transmitting and the receiving purpose, the combination of both transmitting and receiving impulse responses can be treated as a single two-way diffraction filter. The whole process is shown in Fig. 3.4.

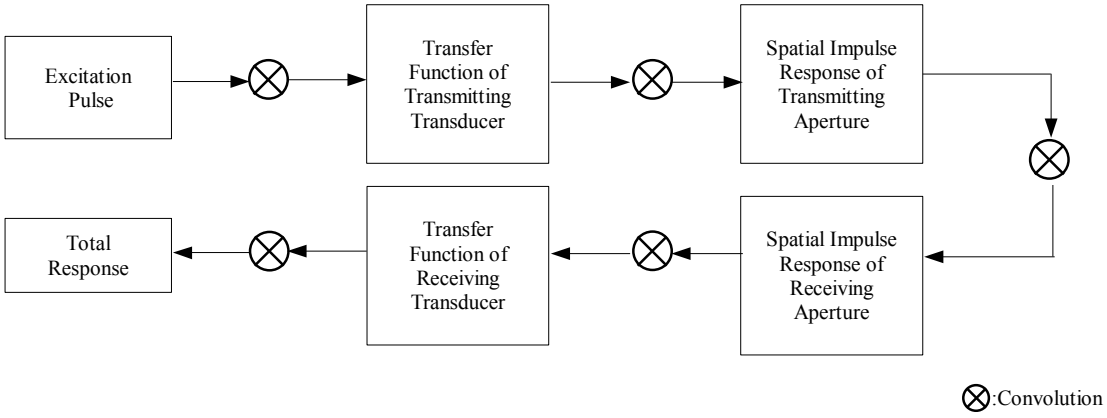


Figure 3.4: Illustration of the total response of an ultrasound transducer element.

The transducer impulse response can be obtained by dividing the transducer surface into small elements and integrating the responses of these elements. Any kind of transducer geometry and excitation can be simulated and both pulse-echo and continuous wave fields can be calculated. Different focusing schemes can be simulated. Both frequency dependent and independent attenuation can be simulated. In our simulation, a single concave focused circular transducer is chosen and the scanning scheme is shown in Fig.1.1. The transducer will be used for both transmitting and receiving. The impulse response of the transducer

and the excitation signal are both simulated as a sinusoidal wave modulated by a Hanning window with certain bandwidth. The simulated tissue is placed in the focal zone of the transducer. When the transducer is excited by a signal it sends the signal out to insonify the tissue and a RF A-line signal is received. This RF signal at time v(t) an be calculated as

$$v(t) = \int_{-\infty}^{+\infty} [\mathrm{T}(\omega)\mathrm{B}(\omega) \sum_{i=1}^{\mathrm{M}} \psi_i(\omega)[\mathrm{A}(\vec{r_i}, \omega)]^2] e^{-\mathrm{i}\omega t} \mathrm{d}\omega, \tag{3.2}$$

where T(ω) is the complex transfer function of ultrasound transducer, T(ω) is the complex spectrum of excitation pulse, $\psi_i(\omega)$ is the scattering strength of i-th scatterer of total M scatterers in the insonified space, and $[\mathrm{A}(\vec{r_i}, \omega)]^2$ is the complex spectrum of the two-way spatial impulse response function (diffraction filter). When the transducer moves along the lateral direction, multiple RF A-lines are obtained. This array of A-lines is then demodulated by using Hilbert transform (see Appendix D) to form a B-scan image. Based on Field II, the author has done some pre-processing and post-precessing work such as the setting of scanning geometry, the simulation of transducer impulse response and excitation signal, setting center frequency and bandwidth, etc. Figure 3.5 shows the whole imaging chain starting with the simulated tissue and ending with the final B-Mode image.

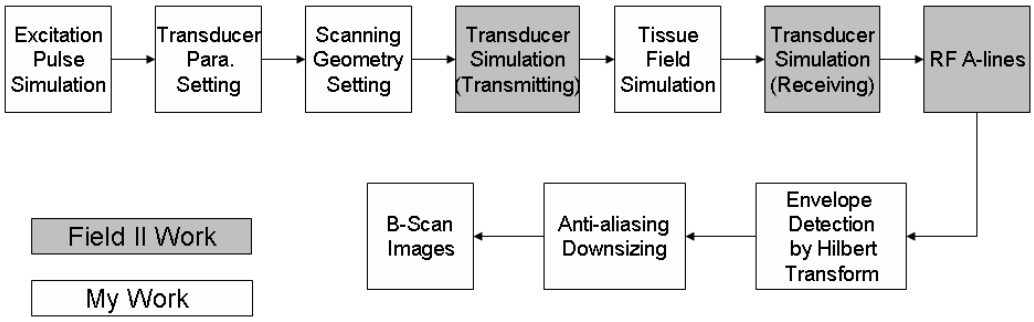


Figure 3.5: Imaging chain of B-Mode imaging simulation.

3.3 First Order Statistical Models of Speckle

In the past years, a lot of first order statistics based speckle models for tissue characterization have been proposed. Many of those models are derived analytically assuming certain tissue scattering properties such as the scatterer number density, randomness of spatial distribution and randomness of scatterer reflectivity. Since we will frequently encounter several such models in this dissertation, a brief introduction to those models is given here. Those models, on the other hand, can be used as a guidance to our tissue parameter setting. Please refer to [15] for more details about those models.

3.3.1 Rayleigh Distribution

The echoed signal from scattering media can be modeled as the sum of backscattered signals from a number of scattering points in the media. The echo signal at a particular instant of time is the sum of signals from the scatterers located in the resolution volume corresponding to that instant of time. The sum is a complex sum due to random phase variations (from random locations of the scatterers) and random amplitude variations (from random reflection strength of each individual scatterer). The net signal is a phasor sum of random amplitude phasors rotating at angular frequency ω_0 if the input signal is a sinusoid with angular frequency of ω_0. All these phasors with different phase and amplitude constitute the so-called *random walk* on the complex plane.

If each scatter of total N scatterers reflects x_i amount of signal, and its phase shift (due to random location) is θ_i, then the phasor sum $\mathbf{x}$ can be written as

$$\mathbf{x} = xe^{j\phi} = \sum_{i=0}^{N-1} x_i e^{j\theta_i}, \tag{3.3}$$

where the three variables x_i, θ_i and N are all random variables. Under the assumptions that the amplitude and phase of i-th phasor are statistically independent, that the amplitude of all phasors are independently and identically distributed (*i.i.d*), that the phase of all the

phasors are uniformly distributed in the range of $(-\pi, +\pi)$, and that the number of scatterers in the insonified region tends to be infinity, it can be shown, using the central limit theorem, that the joint distribution of the real and imaginary part of $\mathbf{x}$ is 2-D Gaussian distribution. The *pdf* of the envelope signal, i.e. the amplitude of the complex phasor, can be evaluated as the well-known Rayleigh distribution:

$$p(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \qquad x \geq 0 \tag{3.4}$$

It is easy to prove that the variance of the random variable x in above is

$$\mathrm{E}\{x^2\} = 2\sigma^2 \tag{3.5}$$

where σ is the single parameter which is related to the total backscattered energy. Another derived parameter of the Rayleigh distribution is the signal-to-noise ratio (SNR):

$$\mathrm{SNR}_{Rayleigh} = \frac{\mathrm{E}\{x\}}{\sqrt{\mathrm{E}\{x^2\} - \mathrm{E}^2\{x\}}} = \frac{\sqrt{\frac{\pi}{2}\sigma}}{\sqrt{2 - \frac{\pi}{2}\sigma}} = 1.914 \tag{3.6}$$

which is a constant and independent of the backscattered energy. It can be used as a criteria for detecting fully formed speckle.

3.3.2 K Distribution

As a limiting case, the Rayleigh distribution is derived under the assumption of fully formed speckle. However, in the real world, the scatterer concentration may not be very high, or there may be some clustering of scatterers. So the breakdown of central limit theorem leads to the deviation from the Rayleigh distribution. K distribution is a non-Rayleigh distribution used by Jakeman [35] to describe the laser speckle statistics with a small number of scatterers. Under certain assumption about the distribution of the cross section of scatterers,

the *pdf* of amplitude signal is given as K distribution below

$$p(x) = 2(\frac{x}{2})^{\alpha}\frac{b^{\alpha+1}}{\Gamma(\alpha)}\mathrm{K}_{\alpha-1}(bx), \qquad x \geq 0, \quad \alpha \geq 1 \tag{3.7}$$

where $\alpha = N(1+\mu)$ is the effective number of scatterers in the resolution volume and N is the actual number of scatterers. μ is a constant which is related to the distribution of the cross section of backscatterers. $\mathrm{K}_{\alpha-1}(\cdot)$ is the modified Bessel function of the second kind of order $\alpha - 1$. Parameter α combines the effects of actual number N of scatterer in the resolution volume, and the severity of amplitude fluctuation. α describes the skewness of the distribution. When α tends to infinity K distribution becomes Rayleigh distribution. If we assume the number of scatterers N obeys negative binomial distribution, which is a generalization of Poisson distribution that allows for the variable mean in Poisson distribution, the same result can be derived without assuming the distribution of the cross section of scatterers. The parameter b is related to both α and the backscattered signal energy as

$$b = 2\sqrt{\frac{\alpha}{\mathrm{E}\{x^2\}}} \tag{3.8}$$

The parameter α can be estimated by using the fourth order central moment $\mathrm{E}\{x^4\}$ and the second order central moment $\mathrm{E}\{x^2\}$ (here we assume the mean of x has been removed). It can be seen that α is simply a function of the (normalized) kurtosis of the signal x:

$$\alpha = \frac{2}{\mathrm{E}\{x^4\}/\mathrm{E}^2\{x^2\} - 2} = \frac{2}{\mathrm{Kurtosis}(x) - 2} \tag{3.9}$$

3.3.3 Rice Distribution

Whenever there is an additional non-random coherent component in the echo signal, which may be due to the unresolved periodically located scatterers or due to strong specular scat-

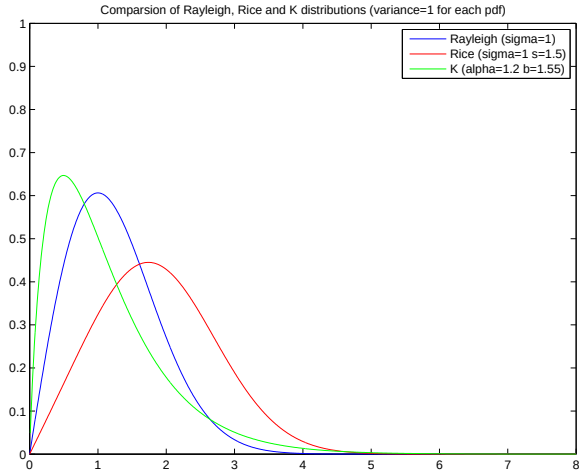


Figure 3.6: Comparison of Rayleigh, Rice and K distributions. For each pdf, the variance is unit. Rayleigh: $\sigma = 1$; Rice: $\sigma = 1$, $s = 1.5$; K: $\alpha = 1.2$, $b = 1.55$.

tering, the phasor model in 3.3 can be modified as,

$$\mathbf{x} = \mathbf{s} + \sum_{i=0}^{N-1} x_i e^{j\theta_i} \tag{3.10}$$

where $\mathbf{s}$ is the non-random coherent component being added. The density function of the envelope signal can be derived as

$$p(x) = \frac{x}{\sigma^2} e^{-\frac{x^2+s^2}{2\sigma^2}} \mathrm{I}_0(\frac{sx}{\sigma^2}), \qquad x \geq 0 \tag{3.11}$$

where I_0 is the modified Bessel function of the first kind of order 0, s^2 is the coherent energy term and $2\sigma^2$ is the diffuse scattering energy. The ratio of s^2 and $2\sigma^2$ is denoted as

$$\gamma = s^2/(2\sigma^2) \tag{3.12}$$

When γ tends to 0, the density function tends to Rayleigh. When γ tends to infinity, the density function tends to Gaussian. The closed form expression of s and σ can be derived from the second and fourth central moments (assume x has been centered) of the backscattered

envelope signal as

$$s^2 = \sqrt{2\mathrm{E}^2\{x^2\} - \mathrm{E}\{x^4\}} \tag{3.13}$$

and

$$\sigma^2 = (\mathrm{E}\{x^2\} - s^2)/2 \tag{3.14}$$

Figure 3.6 compares three *pdf*s of the Rayleigh, Rice and K distributions. Each *pdf* has unit variance. It can be observed that when a *pdf* deviates from Rayleigh to Rice, it tends to be less skewed and closer to a Gaussian. When a *pdf* deviates from Rayleigh to K, it tends to be sharper and more skewed, and its tail becomes longer.

Chapter 4

ICA-based Ultrasound Speckle Texture Analysis and Tissue Characterization

In this chapter, an ICA-based ultrasound speckle texture analysis algorithm was proposed and applied to the simulated B-Mode speckle texture images. Several conventional algorithms were chosen for comparative study. The proposed algorithm and the comparative algorithms were applied to the problems of multi-class tissue classification and multi-class texture based segmentation. By simulating the underlying tissue scatterer fields with a property (or properties) changing gradually, a series of visually similar and gradually changing speckle textures are obtained, possibly representing the pathological changes of soft tissue disease in different stages of severeness. The performance of the proposed algorithm was compared with other algorithms. In many cases, the proposed algorithm shows stronger capability of differentiating textures with very close yet different underlying properties while the comparative algorithms perform worse or fail. We also investigated the behavior of the proposed algorithm for different types of speckle textures. That helps us to understand more about the sophisticated speckle phenomenon from the perspective of higher order statistics.

4.1 Texture Analysis by Multichannel Filtering

Generally, the conventional approaches for ultrasound texture analysis can be classified in five categories: (1) First order statistical approaches, such as those using mean, standard deviation, kurtosis, contrast, entropy or any other derived first order statistical quantities; (2) Second order statistical approaches, such as autocorrelation and co-occurrence matrix methods; (3) Geometrical approaches such as Voronoi tessellation features; (4) Model based approaches such as Markov random field or fractal method; (4) Signal processing approaches (or filtering approaches) such as Fourier power spectrum, discrete cosine transform (DCT), Gabor transform and wavelet transform. Since Gabor transform and wavelet transform are two commonly used ones, a short introduction about these two approaches is given in this section.

In signal processing approaches for texture analysis, a linear filter or filter bank is applied to a texture image and some energy measures are obtained from the filtered image(s). A filter bank is more often used than just a single filter. The basic idea is that the texture can be decomposed into different subbands or subspaces, so that the energy signatures of different textures are different. This approach is similar to how human brain processes texture images. The filters could be spatial domain filters, Fourier domain filters, Gabor filters, wavelet filters, or ICA filters as in our case. Filtering with filter bank is also called multichannel filtering. The general diagram for the multichannel filtering approach is given in Fig. 4.1. In the next section, ICA-based multichannel filtering will be introduced.

The output of multichannel filtering is usually followed by non-linearity and smoothing operations which together are called local energy function. The objective of the local energy function is to estimate the energy in the filter output in a local region. Commonly applied non-linearity are the magnitude $|\cdot|$, the squaring $(\cdot)^2$, and rectified sigmoid $|\tanh(\alpha\cdot)|$. Accurate edge preservation and accurate energy estimation are conflicting goals. For edge localization, high spatial frequency resolution is desired, while for energy estimation, high spatial frequency resolution is desired. These desires need to be balanced in the smoothing

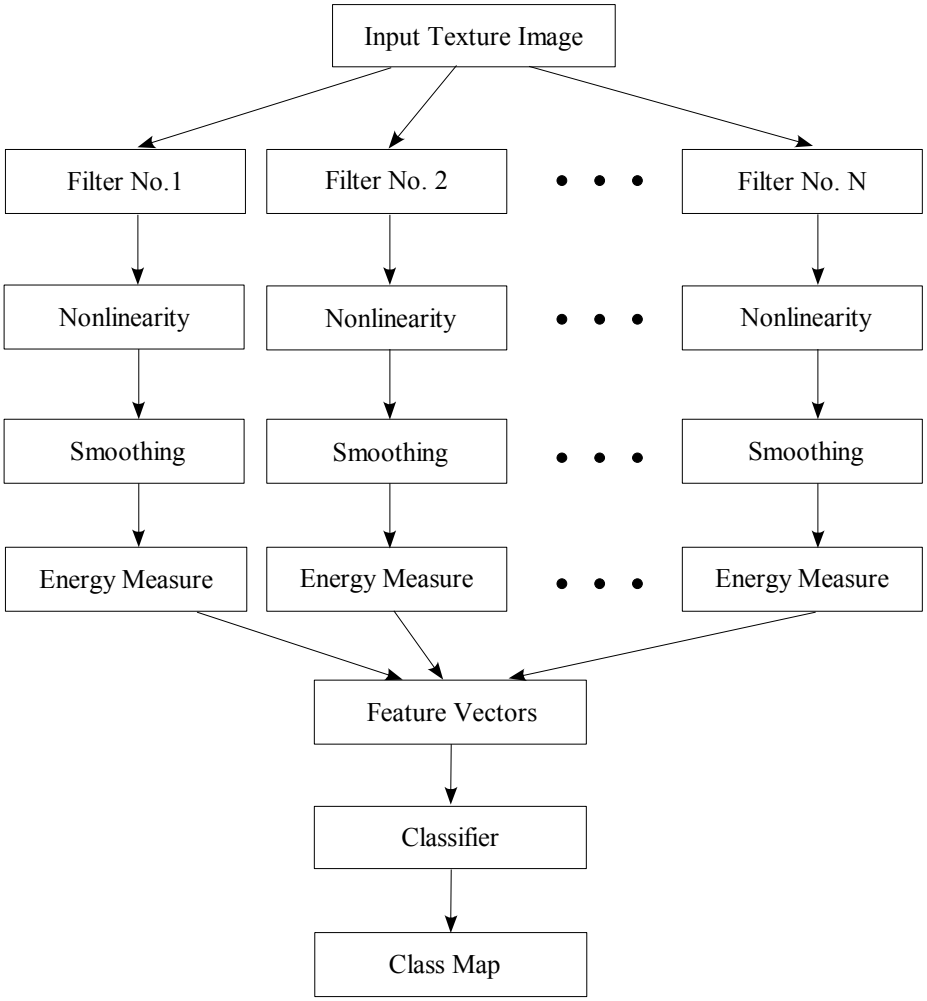


Figure 4.1: A general diagram of multichannel filtering approach for texture analysis.

filter. Usually a Gaussian filter is used as the smoothing filter for a band-limited filter output. The optimal size (i.e., the standard deviation σ) of the Gaussian filter, as suggested in [67], is determined by the center frequency f_0 of the band-limited output image of multichannel filtering, which is given by:

$$\sigma = \frac{1}{2\sqrt{2}f_0} \tag{4.1}$$

Practically, we can choose the form like $\sigma = k/f_0$ where k is a tunable constant which can be determined experimentally.

4.1.1 Gabor Transform

Gabor transform is a windowed short-time Fourier transform when the window is a Gaussian. 2-D Gabor functions are localized with respect to three parameters: spatial location, orientation, and frequency. This is in contrast to Fourier basis functions that are not localized in space, and the basic pixel-by-pixel representation that is not localized in frequency or orientation. The basic 2-D Gabor function for image processing consists of a sinusoidal plane wave of a certain frequency and orientation modulated by a Gaussian envelope. It is given by

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp(-\frac{1}{2}[\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}])\cos[2\pi f_0(x\cos\theta + y\sin\theta)+\phi] \tag{4.2}$$

where f_0 and ϕ are the frequency and phase of the sinusoidal wave. The value σ_x and σ_y are the standard deviation of the Gaussian envelope in the x and y directions respectively . The orientation of the Gabor function is controlled by θ. Figure 4.2 shows a subset of Gabor filters with different combinations of spatial frequencies, phases and orientations.

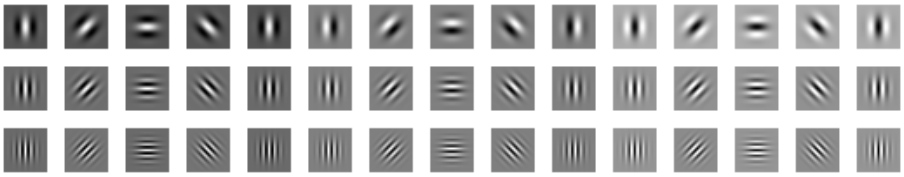


Figure 4.2: A subset of 2-D Gabor filters.

Gabor filters are frequency and orientation selective filters. This can be seen from the Fourier domain analysis of the function. When the θ is 0 and phase ϕ is 0, the Fourier transform of the resulting even-symmetric Gabor function is given by

$$G(u,v) = \exp(-\frac{1}{2}[\frac{(u-f_0)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}]) + \exp(-\frac{1}{2}[\frac{(u+f_0)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}]) \tag{4.3}$$

where $\sigma_u = 1/(2\pi\sigma_x)$, $\sigma_v = 1/(2\pi\sigma_y)$. This function is real-valued and has two lobes in the spatial frequency domain, one centered around and another centered around. For a Gabor filter of a particular orientation, the lobes in the frequency domain are also appropriately rotated.

How to design an optimized Gabor filter bank for a specific texture analysis problem is still a topic for research. In [34], a bank of Gabor filters with dyadic coverage of the radial frequency range and multiple orientations was suggested. This choice was justified by the relation to models for the early vision of mammals as well as the filters' joint optimum resolution in time and frequency. For example, five radial frequencies $\sqrt{2}/2^6$, $\sqrt{2}/2^5$, $\sqrt{2}/2^4$, $\sqrt{2}/2^3$, $\sqrt{2}/2^2$ were suggested in [34] for images of size 256×256. Four orientations $0°, 45°, 90°, 135°$ were used. However, it has been reported in [67] that this method shows rather poor performance for some of the test images. In order to tune the filters to the characteristics of the underlying textures, Bovik *et al.* [3] suggested using narrow band Gabor filters with central frequencies tuned to spectral peaks of textures. That is, for each texture the central frequencies of the corresponding Gabor filters are selected as the frequencies corresponding to the principal spectral peaks of the texture. Some other researchers [87] also addressed the problem of designing an optimized Gabor filter bank for the multi-class texture analysis problem but it has been reported that it is still not a clear winner for different types of images [67].

4.1.2 Wavelet Transform

Wavelet transform is another important multichannel filtering method. Wavelet is based on a single prototype function called mother wavelet $\psi(x)$. The basis functions are obtained by translations $\psi(x + l)$ and dilations (or rescaling) $\psi(x/s)$ of the mother wavelet. 1-D continuous wavelet transform for a given function $f(x)$ can be written as

$$W_f(s, l) = \frac{1}{\sqrt{s}} \int_{t=-\infty}^{\infty} f(t)\psi^*(\frac{t-l}{s})\mathrm{d}t \tag{4.4}$$

For all practical purposes we need a discretized version of the wavelet transform and algorithms which allow a fast and stable computation of the discrete transform. The Discrete Wavelet Transform (DWT) is a special case of the wavelet transform that provides a compact representation of a signal in time and frequency that can be computed efficiently. The 1-D DWT is given as

$$W(j,k) = \sum_{j}^{\infty} \sum_{k}^{\infty} x(k) 2^{-j/2} \psi(2^{-j} n - k) \tag{4.5}$$

where $\psi(t)$ is the mother wavelet function. 2-D DWT, which is frequently used for image processing, can be obtained in a similar way as a 2-D Fourier transform: by first taking the 1-D wavelet transform of all the rows (or all columns) and then the 1-D wavelet transform of the results of previous transform. Figure 4.3 shows some 2-D wavelet basis functions of the Daubechies family of order 10.

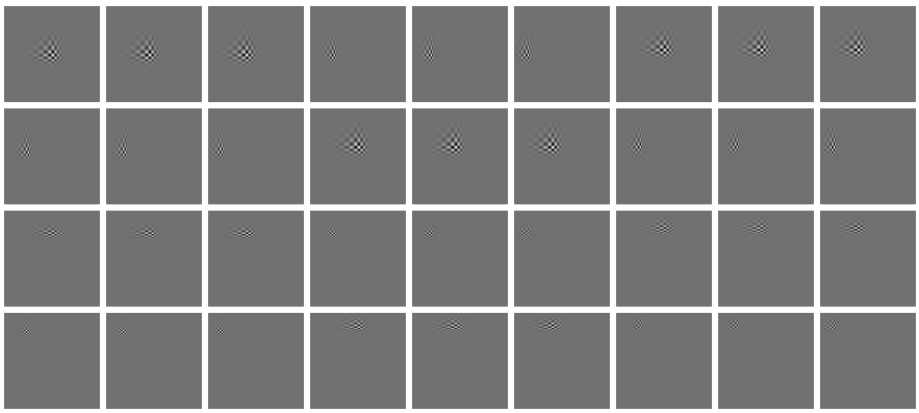


Figure 4.3: A subset of 2-D Daubechies filters of order 10.

DWT corresponds to a critically sampled filter bank with particular filter parameters and subband decompositions. Wavelet transform approaches are consequently filter bank approaches [67]. Mallat [53] used a "standard" wavelet transform for feature extraction, i.e., critically decimated with dyadic subband structure. The work by Chang and Kuo [6] indicates that texture features are most prevalent in intermediate frequency bands, thus the octave band decomposition is not optimal. Hence wavelet packet transform, which basi-

cally is the wavelet transform with subband decompositions not restricted to be dyadic, is used to solve this problem. The discrete wavelet transform and the discrete wavelet packet transform are critically sampled multi-rate filter banks. These types of filter bank typically imply inaccurate texture edge localization. Wavelet frames, which is an overcomplete wavelet representation, is used as a remedy to overcome the problem.

4.2 ICA Filter Bank for Dimensionality Reduction

An image window can usually be modeled as a random field in which each pixel can be treated as a random variable. For an image window $\mathbf{I}$ with total K pixels, $\mathbf{I}$ can be written as a random vector $\mathbf{x} = (x_1, x_2, x_3, \cdots, x_K)$. A Full description of the statistical properties of $\mathbf{x}$ requires the estimation of the joint probability density function

$$p_{\mathbf{x}}(\mathbf{x}) = p_{\mathbf{x}}(x_1, x_2, x_3, \cdots, x_K) = \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}\cdots\frac{\partial}{\partial x_K}F_{\mathbf{x}}(\mathbf{x}) \tag{4.6}$$

where $F_{\mathbf{x}}(\mathbf{x})$ is the cumulative distribution function defined as $F_{\mathbf{x}}(\mathbf{x_0}) = P(\mathbf{x} \leq \mathbf{x_0})$, and $P(\cdot)$ denotes the probability of the event in parenthesis. In fact, the random vector $\mathbf{I}$ may not necessarily represent an image window. It can also be the multi-dimensional features extracted from the window.

When all the random variables in $\mathbf{I}$ are independent then the marginal distributions of those variables can be used to describe $\mathbf{I}$. The 1-D histogram of the image is just a rough representation of this marginal distribution and belongs to first order statistics. The independence condition usually does not hold for real world images. In many situations, the first order statistics is not enough for texture analysis since two textures may have very similar or even identical marginal distributions but totally different visual appearances. Second order statistics based methods, such as the co-occurrence matrix method, only considers the relationship between any pair of pixels. Even for the second order statistics method like the co-occurrence matrix, the computational load is very high. When considering the

relationship among more than two pixels, one encounters the well known *curse of dimensionality*, which is used to describe the problem associated with the feasibility of density estimation in high dimension. When the dimensionality of a multidimensional space becomes large, samples quickly become "lost" in this space, and local neighborhoods become devoid. An acceptable estimation of density requires a number of samples that increases more than exponentially when the number of dimension increases. For example, if $K = 25$ and number of bins L_i used for dimension i is 10, total $\prod_{i=1}^{K} L_i = 10^{25}$ bins are needed to discretely represent the joint distribution of K dimensions. The difficulty is due to not only the heavy computation but also the limited available sample values.

To overcome the difficulty of estimating high dimensional joint density directly, we can decompose the original image into independent features and model each feature by its histogram or marginal distribution. Indeed, the independence of a set of random variables is statistically defined as the equality of the joint distributions of variables and the product of their marginal probability density functions. The Kullback-Leibler (KL) distance (or relative differential entropy) [30], which is defined as a comparison of two probability densities, gives a measure of the similarity between two joint densities. The KL-distance can be defined as

$$\begin{aligned} &KL[p_1(x_1, \cdots, x_K), p_2(x_1, \cdots, x_K)] \\ &\quad = \int_{x_1} \cdots \int_{x_K} p_1(x_1, \cdots, x_K) \log \frac{p_1(x_1, \cdots, x_K)}{p_2(x_1, \cdots, x_K)} \mathrm{d}x_1 \cdots \mathrm{d}x_K \end{aligned} \tag{4.7}$$

If the responses from the filters are independent, then the joint probability density can be factorized as

$$p_j(x_1, \cdots, x_K) = \prod_{i=1}^{K} p_j(x_i), \qquad j = 1, 2 \tag{4.8}$$

Now the KL-distance can be expressed as the sum of the corresponding marginal distribu-

tions, given by the following equation:

$$\begin{aligned} & KL[p_1(x_1, \cdots, x_K), p_2(x_1, \cdots, x_K)] \\ &= \sum_{i=1}^{K} \int_{x_i} p_1(x_i) \log \frac{p_1(x_i)}{p_2(x_i)} \mathrm{d}x_i \\ &= \sum_{i=1}^{K} KL[p_1(x_i), p_2(x_i)] \end{aligned} \tag{4.9}$$

So far the *curse of dimensionality* has been greatly reduced. For example, the number of calculation $\prod_{i=1}^{K} L_i = 10^{25}$ becomes $\sum_{i=1}^{K} L_i = 250$. This method in literature is called independent spectral representation or independent feature representation [52] [54]. Eq. 4.9 establishes the theoretical foundation for texture analysis using ICA. However, the equation may or may not be used directly in any ICA based texture analysis algorithm. In our proposed supervised algorithm, classification is based on a single energy measure of the output of the filter bank instead of the *pdf* estimation for each channel.

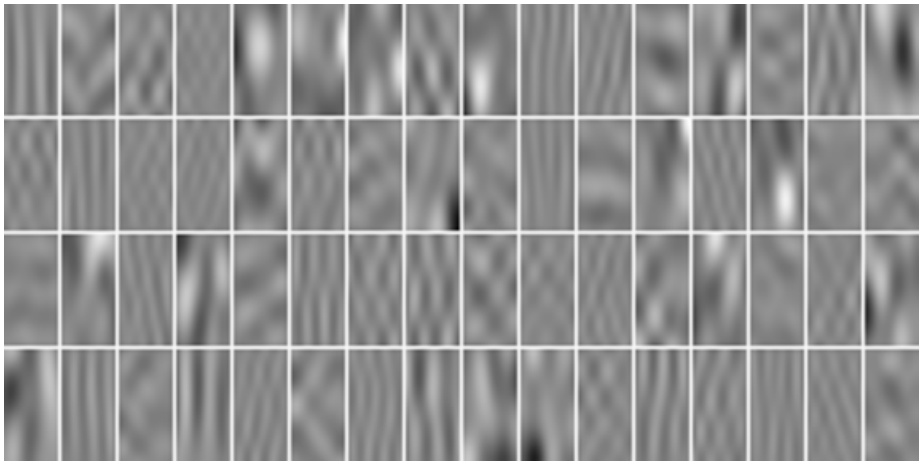


Figure 4.4: A set of ICA basis functions obtained from the training images. Each basis image is in size of 35×17.

In order to obtain ICA filters (or ICA basis functions), a large number of randomly located image windows of a given size are extracted from the training image set. The number of samples is usually decided empirically. In my work, we collected 30,000 samples from 30 training images. Each image window, when reshaped as a 1-D vector, is treated as a

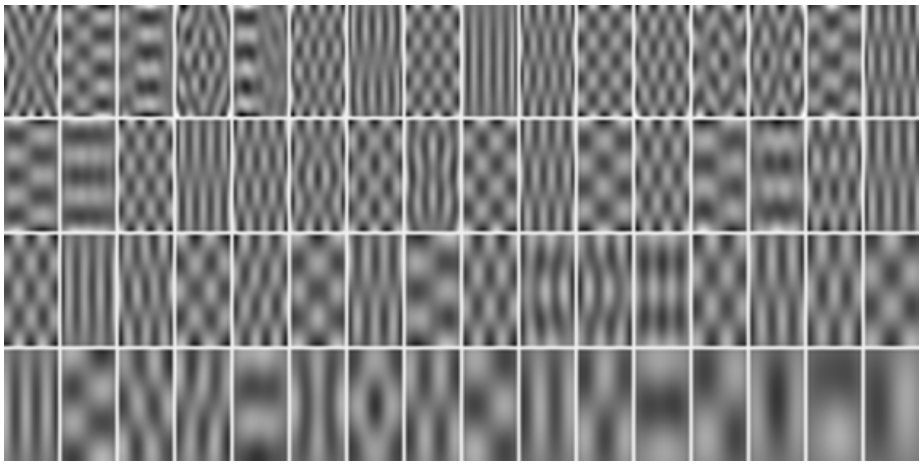


Figure 4.5: A set of PCA basis functions obtained from the training images. Each basis image is in the size of 35×17.

realization of the random vector $\mathbf{x}$ in Eq. 2.6. A proper ICA algorithm is then applied to these observation vectors to obtain the mixing matrix $\mathbf{A}$ (see Eq. 2.6). Each row vector of $\mathbf{W}$, when reshaped as a 2-D image window, can be used as an ICA filter. Each column vector of $\mathbf{A}$, when reshaped as a 2-D image window, can be used as an ICA basis function. Usually we do not differentiate an ICA filter from an ICA basis function since the later one is just a smoothed version of the former one (see Eq. 2.19). Fig. 4.4 shows a set of ICA bases obtained from some simulated B-mode training images.

PCA is usually used as the preprocessing step of an ICA algorithm to reduce the number of basis vectors. The principal components of the sampled image windows, however, can be used as another type of filter bank. PCA is also a technique used to reduce information redundancy but in the sense of second order statistics. Figure 4.5 shows a set of PCA basis functions obtained from the training images.

4.3 Comparison of ICA, PCA, Gabor and Wavelet Filters

Compared with ICA, PCA filtering is a second order statistics based method. It aims to decorrelate the mixed signals so as to obtain the first N most energetic components. If

the signals are mixed uncorrelatedly, PCA can not separate them. PCA is related to linear uncorrelatedness and ICA is related to independence (which is equivalent to non-linear uncorrelatedness). Gabor filtering, wavelet filtering and many other multichannel filtering approaches can be regarded as techniques that are implicitly related to high order statistics.

	2nd Order Statistics	High Order Statistics	Data Adaptation	Orthogonal Bases	Shift and/or Rotation Invariance	Time-Frequency Analysis	Multi-resolution Analysis
ICA		T	T	F	T	F	F
PCA	T	F	T	T		F	F
Gabor		PT	F	F	T	T	T
Wavelet		PT	F	T	F	T	T

Table 4.1: Comparison of multichannel filtering using ICA, PCA, Gabor and wavelet. (T=True, F=False, PT=Partially true or implicitly true)

Gabor bases have different phases that can be changed. Some Gabor bases are similar to bars and others are similar to edges while wavelets bases look like something in between. In Gabor functions, the changes in size and frequency are not independent. The change in size implies a strictly corresponding change in frequency. Compared with Gabor, wavelet has no phase parameters and all the basis functions have the same phase, and hence they all look the same (except with different dilation and translation). In 2-D DWT, there are usually no orientation parameters. The only orientations are horizontal and vertical but in Gabor different orientations are available. However, a wavelet transform usually gives a complete set of orthogonal bases while a Gabor transform does not.

Gabor filters are similar to ICA filters which give some image features with different shape, phase and orientation. But for Gabor transform, it is still an open problem as to what set of values should be chosen for the parameters to obtain optimal representation of the data. Gabor bases and wavelet bases are both general fixed bases with some analytical forms. Unlike ICA bases, they usually are not estimated from data so they may not be the best representation for a specific problem. On the other hand, ICA bases are estimated

from data directly so they may be more data adaptive. Wavelet transform gives a complete set of orthogonal base of 1-D space while Gabor does not. The drawback of ICA bases is that they can not be used for general data set. Like Gabor, ICA also does not give a complete optimal orthogonal basis set. ICA algorithms also suffer from the local minimum problem just like many artificial neutral network algorithms. Gabor and wavelet transforms are localized in both time and frequency domains. They are both multi-resolution analysis, which means that an image can be analyzed at different resolutions or frequencies. ICA basis functions, however, are usually estimated from the samples of same windows size so that they lack the "zoom-in" and "zoom-out" nature of wavelet transform. The discussion above is summarized in Table 4.1.

4.4 Using Support Vector Machines (SVM) as Classifier

The performance of texture analysis by multichannel filtering is greatly affected by the capability of the classifier chosen. In this study, Support Vector Machines (SVM) [13], which is a state-of-the-art pattern recognition technique developed in recent years, is used as the classifier. SVM is based on the statistical learning theory to minimize the classification errors of the training data and unknown data. The goal of SVM is to find an optimal hyperplane that separates clusters of vectors in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other side of the plane. The vectors near the hyperplane are the support vectors, which are actually the linear combination of some input feature vectors. A good separation is achieved by the hyperplane that has the largest distance to the neighboring data points of both classes, since in general the larger the margin the better the generalization error of the classifier.

SVM solves the non-linear separation problem by mapping original input space to higher dimension space, i.e., the so-called feature space in SVM literature. The mapping

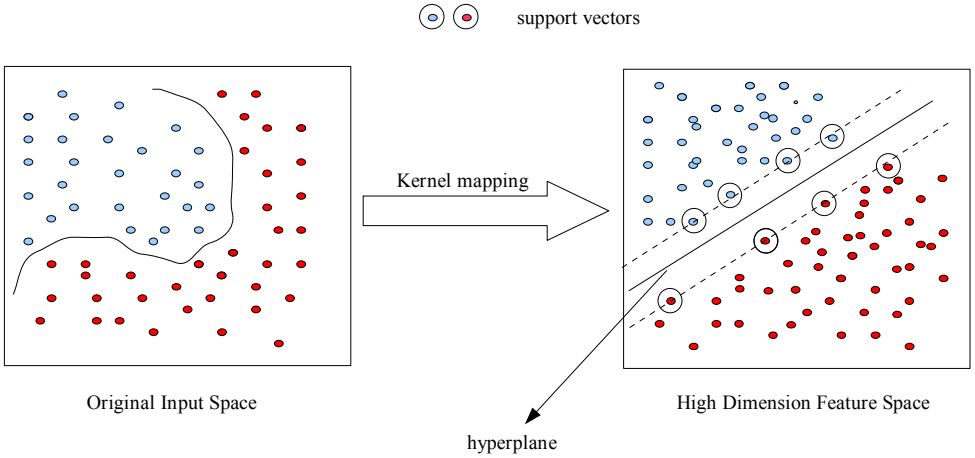


Figure 4.6: An illustration of the kernel mapping of SVM. The nonlinear separable feature vectors in original input space are mapped to the higher dimensional linear separable feature space. A hyperplane is obtained with support vectors on each side of it.

is implemented implicitly by a kernel function, such that linear decision boundaries constructed in the high dimensional feature space correspond to nonlinear decision boundaries in the input space. In feature space, an optimal separating hyperplane is constructed by solving an optimization problem which can be converted to a convex quadratic program subject to some linear constraints. Figure 4.6 illustrates kernel mapping of SVM and the hyperplane defined by support vectors.

Mathematically, the basic problem of SVM can be described as follows. Given a training set of instance-label pairs $(\mathbf{x}_i, y_i), i = 1, 2, \cdots, l$ where $\mathbf{x}_i \in \mathrm{R}^n$ and $\mathbf{y} \in \{-1, 1\}^l$, the support vector machines require the solution of the following optimization problem:

$$\min_{\mathbf{w},b,\xi} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i$$

$$\text{subject to} \qquad y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0. \tag{4.10}$$

Here $\mathbf{w}$ is a vector of hyperplane coefficients and b is a bias term. The training vectors $\mathbf{x}_i$ are mapped into a higher dimensional feature space by the function ϕ. Then SVM finds a

linear separating hyperplane with the maximal margin in this higher dimensional space. C (> 0) is the penalty parameter of the error term. Furthermore, $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ is called the kernel function. Most frequently used kernels include:

- linear:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j. \tag{4.11}$$

- polynomial:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d, \ \gamma > 0. \tag{4.12}$$

- radial basis function (RBF):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2), \ \gamma > 0 \tag{4.13}$$

Here γ, r and d are kernel parameters.

SVM is a classifier with good generalization and is among the best performers for a number of classification problems. SVM is a close cousin to traditional multilayer neural networks. In training multilayer neural network classifiers, network weights are usually corrected so that the sum-of-squares error between the network outputs and the desired outputs is minimized. But because the decision boundaries between classes acquired by training are not directly determined, classification performance for the unknown data, i.e., the generalization ability, depends on the training method. And it degrades greatly when the number of training data is small and there is no class overlap. When the dimension of an input vector become relatively high (for example, larger than 10), the multilayer neural network performs poorly and may encounter serious convergence problem. SVM has been successfully applied to the segmentation of ultrasound image using the pixel grey level values [46]. Our experience has shown that SVM outperforms the traditional 3-layer feed-forward neural network with back-propagation (FF-BP). Among the three kernels mentioned above, the RBF kernel usually seems to be the best performer.

4.5 Proposed ICA-based Algorithm

The proposed algorithm consists of three stages: The first stage is learning ICA basis functions from the training image set; The second stage is obtaining feature vectors by multichannel filtering, reducing the dimension of feature vectors by ICA filter selection and training the SVM model; The third stage conducts texture classification or segmentation using the trained SVM model. Some selected conventional algorithms for comparison study are also discussed.

4.5.1 Stage I: Learning ICA Basis Functions from Training Images

The diagram for this stage is shown in Fig. 4.7. The input images for training are obtained by simulation, as discussed in Chapter 3, and have been properly envelope-detected and downsampled. In step 1, the average speckle size is estimated by taking the full-width-at-half-maximum (FWHM) of the auto-covariance function (ACVF) of the images [79]. ACVF is defined as

$$\mathrm{ACVF}(x,z) = \iint_{ROI} \{\mathrm{I}(\xi+x,\zeta+z)-\mathrm{I}_0\}\{\mathrm{I}(\xi,\zeta)-\mathrm{I}_0\}\mathrm{d}\xi\mathrm{d}\zeta \tag{4.14}$$

where I is the envelope image of a ROI (region-of-interest) and I_0 is the mean value of I. x, ξ are lateral coordinates and z, ζ are axial coordinates. This speckle size estimation should be conducted on images of all tissue types and averaged out, since the speckle size not only depends on the transducer characteristics but also varies with the scatterer number density as well [63]. The window size for ICA sampling is critical to the performance of the algorithm. Too small window size causes the algorithm to fail to catch the local texture variations as tissue signatures while too large window size causes very heavy computational burden for an ICA algorithm and sometimes makes it difficult to converge. Our experience shows that if the size of sampling window is set to be about 2 to 4 times the averaged speckle size in both the axial and lateral directions, then a good trade-off between performance and

computation burden can be achieved. The exact size of sampling window can be found in Section 4.7.

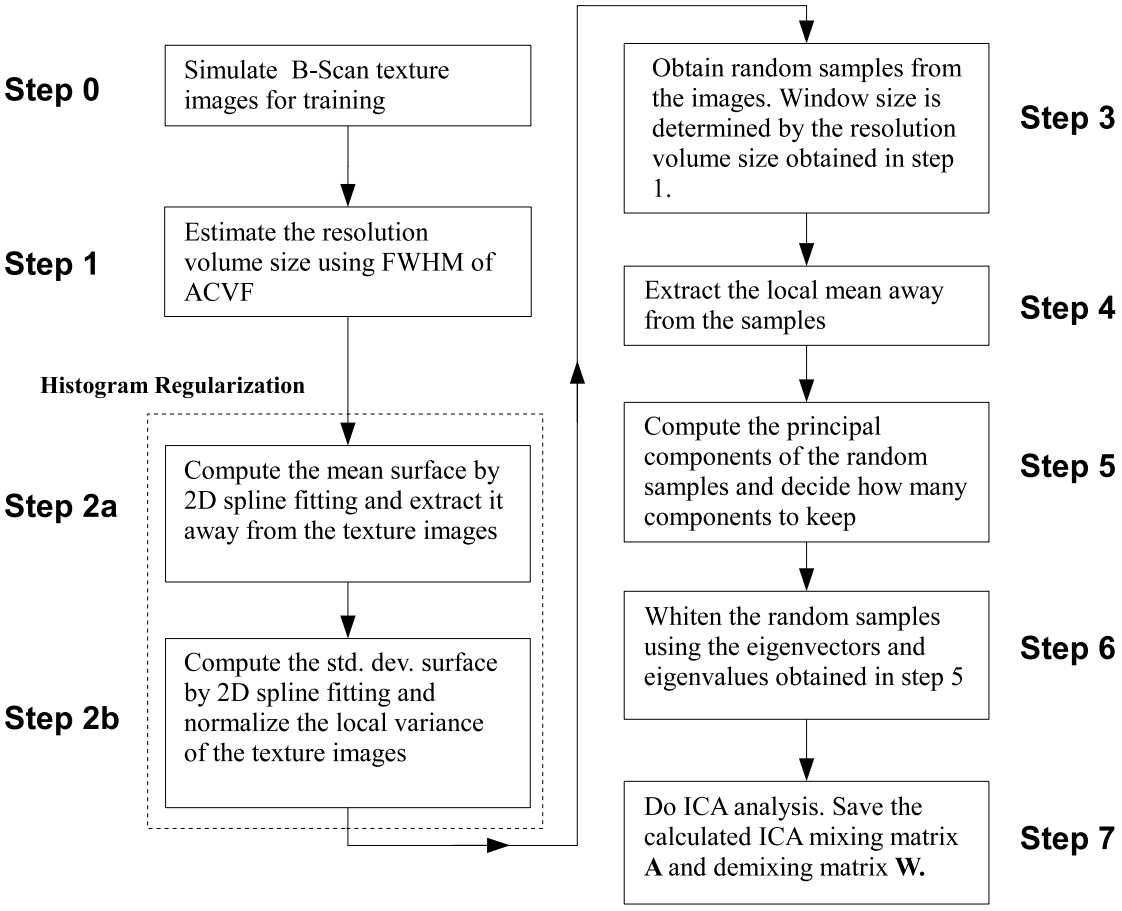


Figure 4.7: Diagram of proposed algorithm (Stage I): Learning ICA basis from training images.

Since the spatial impulse response of ultrasound transducer is a function of axial distance, the ultrasound image appears to be non-stationary along the axial direction. For a focused transducer, the axial beam profile shows a peak at the focal distance, and then attenuated fast when axial distance increases. Usually tissues are imaged at the focal zone in order to obtain high resolution. Only if the size of the tissue at the focal zone is small enough compared to focal zone size, the image obtained can be treated to be quasi-stationary. Usually if the tissue size is not small enough, the B-mode image can be divided into several blocks at different axial depths and then each block can be treated as quasi-stationary. Since

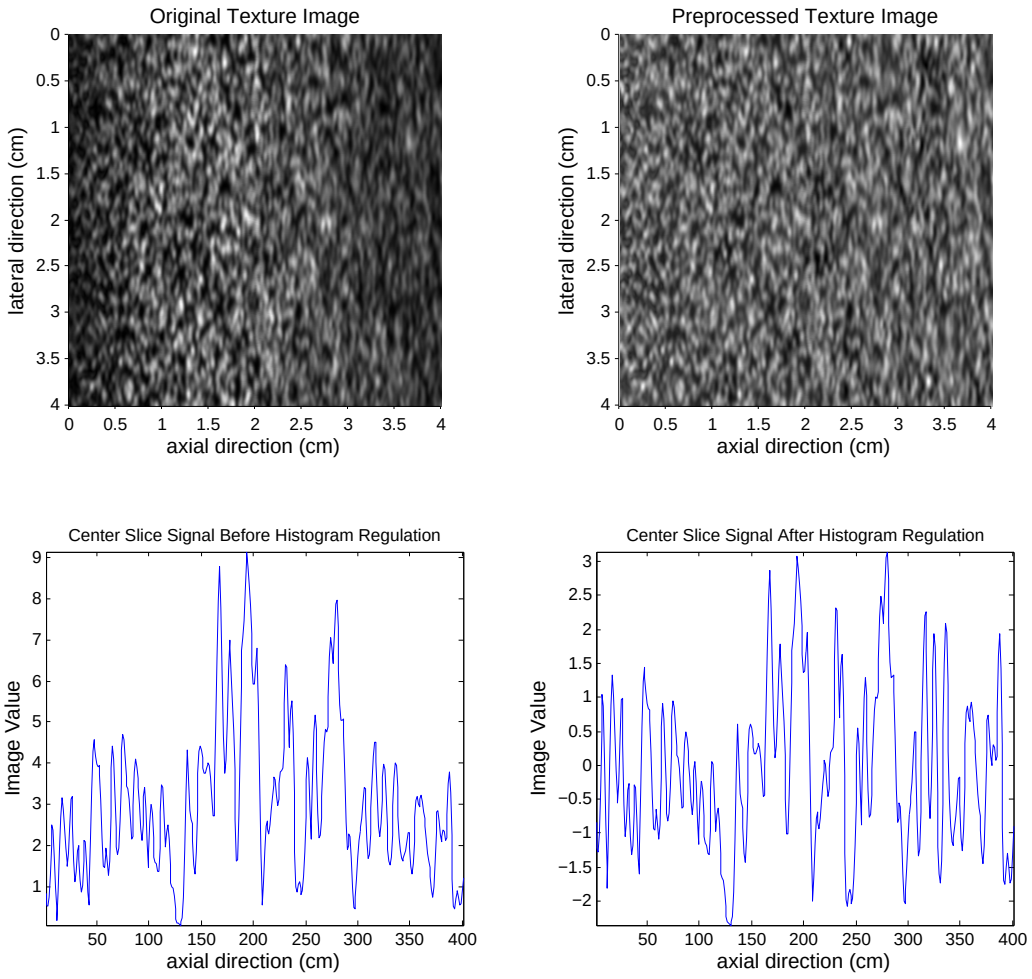


Figure 4.8: B-mode images before and after histogram regulation, and their corresponding center slice signals.

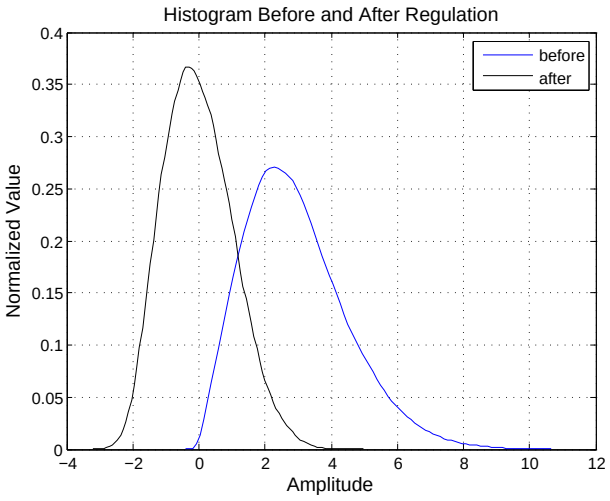


Figure 4.9: Histograms of the B-mode images before and after histogram regulation.

most classical ICA algorithms can not handle well non-stationary mixed signals [30], non-stationary B-mode images will cause an ICA learning algorithm to become very unstable and sometimes hard to converge. To overcome this difficulty, a method is proposed here to reduce the non-stationarity by regularizing the histogram so that the mean is removed and the variance is normalized, both locally and globally. One should notice that only wide-sense non-stationarity is considered here, i.e., only the mean and variance will be dealt with. To remove the mean away, 2-D spline is used to fit the sampled local mean values. The sampled local mean values are calculated from a large number of image windows. Then the fitted mean surface is subtracted from the image. To normalize the variance, 2-D spline is used to fit the sampled local standard deviation values from many image windows of the mean removed image. Then the fitted standard deviation surface is divided by the mean-extracted image so that the variance is normalized. The 2-D spline fitting algorithm used in this study is called least-squares spline approximation using B-form spline of order k with a given sequence of knots. Matlab function *spap2* was called in my program. In order to avoid boundary problems, a processed 500×500 image was cropped to the 400×400 center part of it.

This histogram regularization process makes the images quasi-stationary. A byproduct of this histogram regularization step is that the first order statistics (up to second order moments) of different texture images become approximately the same, which would be helpful for us to validate the performance of some multichannel filtering techniques such as ICA or PCA which relies on second or higher order statistics only. In many situations, the significance of mean and variance values would be trivial because they may be imaging system dependent and can not provide further information about tissue microstructure so they need to be removed. After the histogram regularization, the visual patterns of the texture images keep the same. This indicates that the 1-D histogram regularization does not significantly change the second or higher order statistics. Figure 4.8 shows the B-mode images before and after the histogram regularization, and their corresponding central slice

signals. Figure 4.9 shows the histograms before and after the regularization.

After the preprocessing step of histogram regularization, random image samples are collected from the the training images, with sampling window size set as two times of the FWHM of ACVF on both axial and lateral directions. The local mean values are extracted from those samples and then principal components are computed. The first N principal components are kept to guarantee that at least 95% of signal energy is kept. The percentage of energy to keep is usually based on experience and experiments. Usually 95% to 99% would be a reasonable range. The random samples are then whitened using the eigenvectors and eigenvalues obtained in the previous step. A proper ICA algorithm (FastICA C is used in this study) is then applied to the whitened signals to obtain the mixing matrix A and demixing matrix W.

4.5.2 Stage II: Feature Extraction, Filter Selection and SVM Training

The diagram for this stage is shown in Fig. 4.10. The same set of training images used in Stage I are also used in this stage. First, the images go through the same preprocessing step of histogram regularization, and then multichannel filtering using the ICA basis functions are applied. If the number of images is M, and the number of ICA filters is N, then total $M \times N$ filtered images are obtained. Then the nonlinearity $|\cdot|$ (taking the absolute value) is applied to all these images. For each filtered image, the variance σ of the smoothing Gaussian filter is determined by the peak frequency of its Fourier spectrum (see Eq. 4.1), and then the image is smoothed by the Gaussian filter. The smoothed images in step 6 are called feature images. A feature vector of dimension N is then obtained by taking the feature image pixel values of the same sampling location across total N feature images.

Usually each dimension of the feature vectors is called an attribute. In step 8, each attribute of the feature vectors is scaled to the same range such as [-1, 1]. Scaling before applying SVM is very important [26]. The main advantage is to prevent attributes in greater numeric ranges dominating those in smaller numeric ranges. Another advantage is to avoid

numerical difficulties during the calculation. Because kernel values usually depend on the inner products of feature vectors, large attribute values might cause numerical problems. Exactly the same range must be used to scale both the training and testing data.

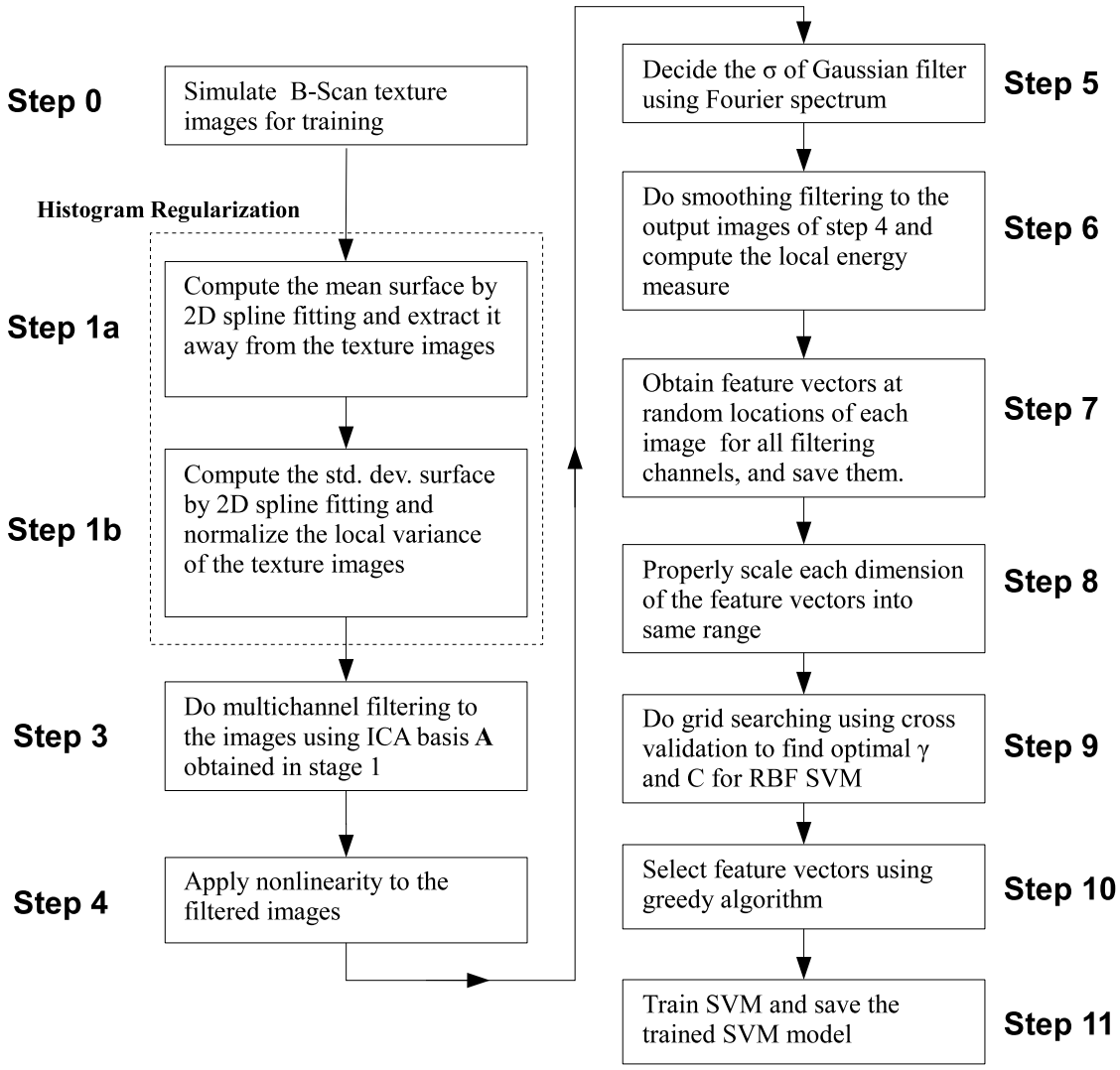


Figure 4.10: Diagram of proposed algorithm (Stage II): Feature extraction, selection and SVM training.

In step 9, a grid-search with cross-validation was conducted to obtain optimal C and γ, the two parameters of SVM with the RBF kernel. It is not known beforehand which C and γ are the best for one problem; consequently some kind of model selection (parameter

search) must be done. The goal is to identify good (C, γ) so that the classifier can accurately predict unknown data (i.e., testing data). Note that it may not be useful to achieve high training accuracy (i.e., classifiers accurately predict training data whose class labels are indeed known). Therefore, a common way is to separate training data into two parts of which one is considered unknown in training the classifier. Then the prediction accuracy on this set can more precisely reflect the performance on classifying unknown data. An improved version of this procedure is cross-validation. In ν-fold cross-validation, we first divide the training set into ν subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining $\nu - 1$ subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. The cross-validation procedure can prevent the overfitting problem. A simple grid-search on C and γ using cross-validation is conducted. Basically pairs of (C, γ) are tried and the one with the best cross-validation accuracy is picked. Trying exponentially growing sequences of C is a practical method to identify good parameters (for example, $C = 2^{-5}, 2^{-3}, \cdots, 2^{15}, \gamma = 2^{-15}, 2^{-13}, \cdots, 2^{3}$).

After step 9, the dimension of feature vectors may still be too high. For example, if we assume that the ICA sampling window size is 35×17, then the dimension of 595 (= 35×17) can be reduced to about 64 after principal components selection. However, dimension 64 is still challenging to any classifier, hence needs to be further reduced. Another concern is that not all the ICA filters can provide the best discriminating capability. To further reduce the dimension of feature vectors, a greedy algorithm is used here. The basic idea is that a subset of the ICA filters can be selected so that the overall performance is close to that using all the filters. A greedy algorithm is used to select a small set of ICA filters to improve the performance [52]. First, all the training samples are bootstrapped into two sets: training set and validation set. The basic idea is to choose filters one by one so that the performance estimated on the validation set is maximized. The available samples are divided into L training-validation pairs, $T^{(1)}, V^{(1)}, T^{(2)}, V^{(2)} \cdots, T^{(L)}, V^{(L)}$, where $T^{(i)} \bigcup V^{(i)}$

is full set and $T^{(i)} \bigcap V^{(i)} = \phi$. We estimate the recognition performance on $V^{(i)}$ by using the training samples in $T^{(i)}$ and the filter in C, denoted by $F(V^{(i)}|T^{(i)}, C)$. The algorithm can be summarized as follows, where $\mathbf{W}$ is the set of candidate ICA filters, and C is the chosen subset of the ICA filters so far and ϵ is a threshold as convergence condition.

$\mathbf{W} = \{\overrightarrow{W}_1, \cdots, \overrightarrow{W}_K\}, \quad C = \phi$
Randomly generate L training-validation pairs $T^{(1)}, V^{(1)}, T^{(2)}, V^{(2)} \cdots, T^{(L)}, V^{(L)}$
repeat
    for each filter $\overrightarrow{W}_\alpha$ in $\mathbf{W}$
    calculate $F(C \bigcup \overrightarrow{W}_\alpha) = (1/L) \sum_{i=1}^{L} F(V^{(i)}|T^{(i)}, C \bigcup \overrightarrow{W}_\alpha)$
    $\alpha^* = \max_\alpha F(C \bigcup \overrightarrow{W}_\alpha)$
    $E = F(C \bigcup \overrightarrow{W}_{\alpha^*}) - F(C)$
    if $E > \epsilon$ then $C = C \bigcup \overrightarrow{W}_{\alpha^*}, \quad \mathbf{W} = \mathbf{W} \setminus \overrightarrow{W}_{\alpha^*}$
until $E \leq \epsilon$

Table 4.2: Greedy algorithm for filter selection

Such kind of greedy algorithm may not guarantee the final solution to be globally optimal. However, usually the sub-optimal solution obtained by a greedy algorithm is practically good enough. Compared with some exhaustive algorithms such as dynamic programming, the computational burden using the greedy algorithm is greatly reduced.

4.5.3 Stage III: Classification and Segmentation

In step 0 of Stage III, texture images which are independent of the training images are generated to test the performance of the proposed algorithm. For classification, multiple images of different tissue classes are used as input. For segmentation, a single two- or multiple-component image is used as input. From step 1 to step 8, an input image passes the same processing steps as those in stage 2. In step 9, the trained SVM model obtained in stage 2 is used to classify each pixel of the input image. Usually a running window is used to go through the whole image. The class label of the center pixel of the window is obtained. For a segmentation job, smaller window size would be helpful to obtain sharper

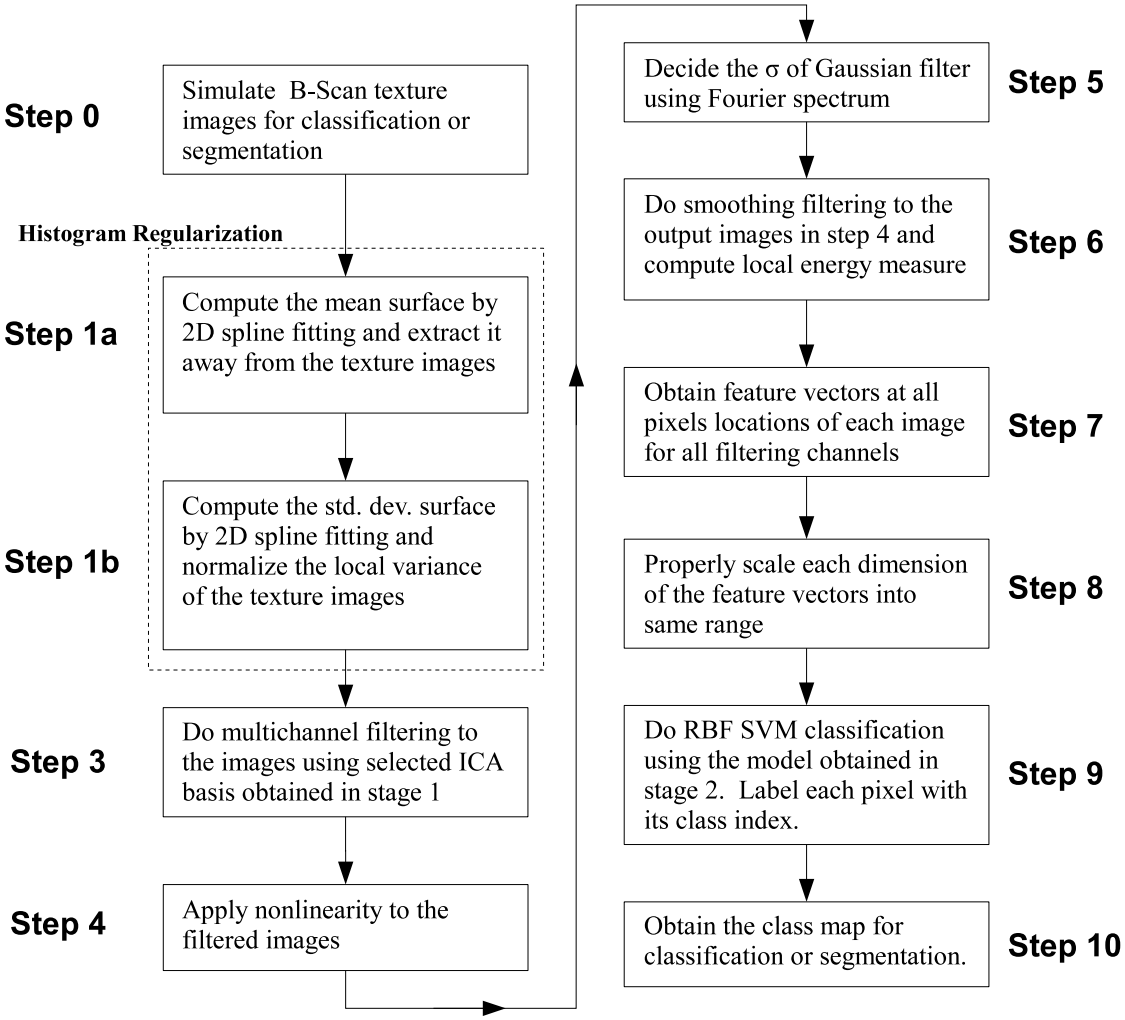


Figure 4.11: Diagram of proposed algorithm (Stage III): Classification or segmentation

segmentation boundary but the overall pixel classification rate may be reduced and the class map looks noisy. Larger window size will be helpful to obtain smoother class map but the object boundary will look soft and fuzzy hence some trade-offs need to be made.

4.6 Algorithms for Comparison

There are so many algorithms for texture analysis developed in the last several decades, and each algorithm may have its own advantage in a specific application. It is almost impossible to compare all of those algorithms thoroughly with our proposed algorithm. For selecting

appropriate algorithms for comparison, the following criteria have been established: (1) The selected algorithms should include classical and state-of-the-art ones, especially those being used in the field of ultrasound texture analysis; (2) The selected algorithms should include first, second and high order statistics related ones; (3) Classical multichannel filtering algorithms need to be considered because they are on the same line of approach as the proposed ICA-based algorithm; Non-filtering methods also need to be considered. (4) The realization of the selected algorithms should be straightforward. Highly sophisticated algorithms with a lot of variant forms and parameter tweaks should be avoided; (5) Algorithms combining two or more major techniques, for example, Gabor combined with ICA, should be avoided. Based on the above criteria, the following algorithms are selected for the comparison study:

- *SNR-ALPHA*. This method uses two first order statistics parameters, SNR defined in Eq. 3.6 and α defined in Eq. 3.9, as the features for classification. We rewrite those two definitions here as:

$$\mathrm{SNR}_{Rayleigh} = \frac{\mathrm{E}\{x\}}{\sqrt{\mathrm{E}\{x^2\} - \mathrm{E}^2\{x\}}} \tag{4.15}$$

and

$$\alpha = \frac{2}{\mathrm{E}\{x^4\}/\mathrm{E}^2\{x^2\} - 2} \tag{4.16}$$

Each parameter is estimated from the same size of sampling window as that for ICA basis learning. The workflow of this algorithm is similar to Stage II (Fig. 4.10) and Stage III (Fig. 4.11) of the proposed algorithm. However, an input image does not go through the histogram regularization step and the filter selection step. The reason to pick the two parameters above as features is that SNR is the definitive quantity to describe Rayleigh distribution while α (effective number of scatterers in the resolution volume) is a quantity to describe how far away the signal is deviated from Rayleigh distribution to K-distribution.

- *SNR-GAMMA*. This method uses two first order statistics parameters, SNR (see Eq. 4.15) and γ (the ratio of coherent and diffuse energy component) defined in Eq. 3.12, as the features for classification. We rewrite γ here as:

$$\gamma = s^2/(2\sigma^2) \tag{4.17}$$

  The reason to pick γ as a feature is because it is a quantity to describe how far away the signal is deviated from Rayleigh distribution to Rice distribution. The workflow of this algorithm is similar to SNR-ALPHA mentioned above.

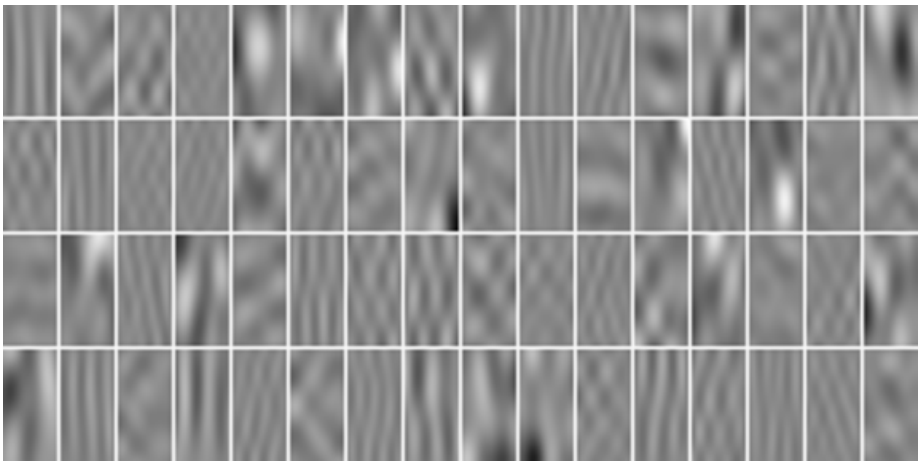
- *FWHM-ACVF*. This method uses two parameters, the full-width-at-half-maximum (FWHM) along both axial and lateral directions of the auto-covariance function (ACVF) (defined in Eq. 4.14) of the image window, as the features for classification. We rewrite the definition of ACVF here as:

$$\mathrm{ACVF}(x,z) = \iint_{ROI} \{\mathrm{I}(\xi+x,\zeta+z) - \mathrm{I}_0\}\{\mathrm{I}(\xi,\zeta) - \mathrm{I}_0\}\mathrm{d}\xi\mathrm{d}\zeta \tag{4.18}$$

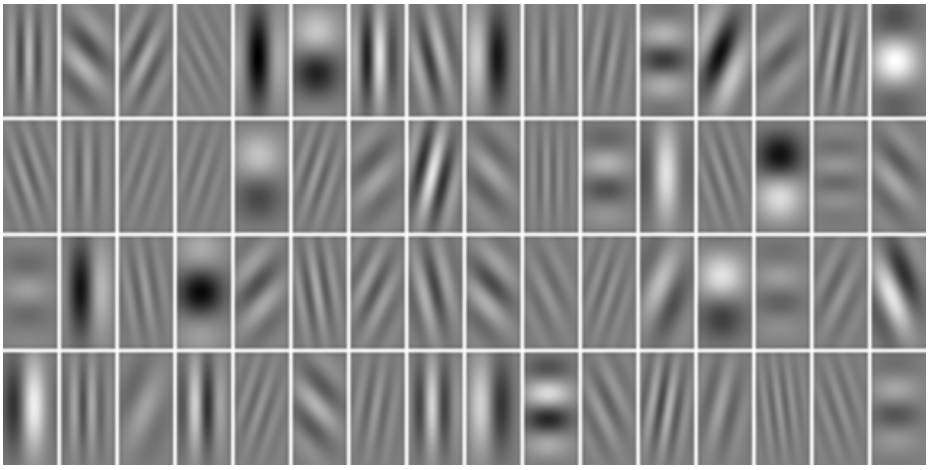
  ACVF is estimated from the same size of window as that for ICA basis learning. FWHM of ACVF is a classical state-of-the-art method for ultrasound texture analysis. It is a typical second order statistics method which considers only the pixel correlation along axial and lateral directions. The workflow of this algorithm is very similar to that of SNR-ALPHA and SNR-GAMMA. However, an input image needs to go through the histogram regularization step.

- *PCA-based Multichannel Filtering*. It is easy to have a head-to-head comparison between this method with the ICA-based method. The comparison will directly show the performance difference of the second order vs. the high order statistical method. The workflow of this algorithm is very similar to that of the proposed algorithm: In the diagram of Stage I (Fig. 4.7), simply ignore step 7 since PCA filters have already

been obtained in step 6. In the diagram of Stage II (Fig.4.10) and III (Fig. 4.11), simply use the PCA filters to replace the ICA filters, while all other conditions are kept exactly the same.



(a) A set of ICA filters



(b) The corresponding Gabor filters

Figure 4.12: A set of ICA filters and their corresponding Gabor filters.

- *Gabor-based Multichannel Filtering*. In this method, Gabor filters are used for multichannel filtering with the workflow similar to the proposed algorithm. Gabor filtering is a classical method for texture analysis. Gabor filters have very similar properties as the ICA filters (see 4.3) and at the same time Gabor analysis has a similar time-frequency analysis property as wavelet analysis (that is why Gabor functions are sometimes called Gabor wavelet). So Gabor, instead of wavelet, is chosen as a rep-

resentative of multichannel filtering techniques for comparison, instead of choosing wavelet. Since there is no uniformly agreed method for selecting an optimal set of Gabor filters, an alternative method is used in this study to generate Gabor filters. The Gabor filters' central frequencies are tuned to the spectral peaks of the ICA filters. That is, for each ICA filter the central frequency of the corresponding Gabor filter is selected as the Fourier spectral peak of that ICA filter. The orientation and phase can also be decided by the spectral peak. The only parameters remaining to be determined are the size of the filter, i.e., standard deviation σ_x and σ_y of the Gaussian envelope (see Eq. 4.2). σ_x and σ_y are determined by searching in a range of values and selecting the best candidate so as to minimize the mean square error between the Gabor filter and the corresponding ICA filter. Figure 4.12 shows a set of ICA filters and their corresponding Gabor filters.

4.7 Experimental Settings

4.7.1 Tissue Sets Used in the Experiments

A total of 4 sets of tissues are simulated. Each set contains 6 tissue types (scatterer field types) with a gradually changing parameter. Each set of tissues have been generated to match the speckle properties predicted by the classical first order statistics models introduced in Chapter 3. Since the searching space of tissue parameter combinations is too large, we only consider those sub-regions in which the speckle approaches the limiting Rayleigh region from different non-Rayleigh regions. For example, a series of tissues (with labels from 0 to 5) may have corresponding speckles approaching Rayleigh distribution from the Rice distribution with increasingly changing spatial regularity, or approaching the Rayleigh distribution from the K distribution with decreasingly changing effective scatterer number density. Table 4.3 gives a full description of the 4 simulated tissue sets used in our experiments.

A total of 4 grand experiments were conducted. Each experiment corresponds to one tissue set (e.g., Exp. No. 1 uses tissue set No. 1). For each tissue type in a tissue set, 5 training images and 1 testing image were simulated so a total of 30 training images and 6 testing images were generated. Several two-component testing images with different combinations of 2 out of 6 tissue types as background and lesion regions were generated for segmentation task. Each two-component image has a blob-shape lesion region embedded inside the background. The proposed ICA algorithm plus five other comparative algorithms were applied to those simulated images.

Tissue set T1 was used in experiment No. 1. In T1, the scatterer number density and scatterer reflectivity are kept the same for all 6 tissue types while the scatterer spatial randomness is controlled by the Poisson disk process with the minimum inter-scatterer distance d_{min} changing from 4.0mm to 1.5mm in the step of 0.5mm. The purpose of this experiment is testing texture classification and segmentation for tissues with high scatterer number density and with spatial randomness gradually approaching fully random case (small d_{min}) from very regular case (large d_{min}).

Tissue set T2 was used in experiment No. 2. In T2, the scatterer reflectivity is fixed and the scatterer number density ranges from $1.44/mm^2$ to $4.84/mm^2$. The scatterer spatial randomness is controlled by the Poisson disk process with a minimum inter-scatterer distance d_{min} ranging from 0.5mm to 0.27mm. Here the value of d_{min} is roughly set as $0.6/\sqrt{SND}$ where SND is the scatterer number density. The purpose of this experiment is testing texture classification and segmentation for tissues with scatterer density slightly deviating from very high density to moderately high density, while the scatterer spatial distribution is kept to be slightly regular and correlated to the scatterer number density.

Tissue set T3 was used in experiment No. 3. In T3, the scatterer reflectivity is kept fixed and the scatterer number density ranges from $1.44/mm^2$ to $4.84/mm^2$. The scatterer spatial locations are kept fully random by using the homogenous Poisson process. The purpose of this experiment is testing texture classification of tissues deviating from very high scatterer

Tisue Set Code Name	Class Index	Tissue Code Name	Total Scatterer Number	Est. Scatterer Number Per Resolution Cell (unit: 1/mm^2)	Scatterer Spatial Randomness (unit: 0.1mm)
T1	0	T1-0	10000	12.00	Poisson Disk (d_{min}=4.0)
	1	T1-1	10000	12.00	Poisson Disk (d_{min}=3.5)
	2	T1-2	10000	12.00	Poisson Disk (d_{min}=3.0)
	3	T1-3	10000	12.00	Poisson Disk (d_{min}=2.5)
	4	T1-4	10000	12.00	Poisson Disk (d_{min}=2.0)
	5	T1-5	10000	12.00	Poisson Disk (d_{min}=1.5)
T2	0	T2-0	3600	4.32	Poisson Disk (d_{min}=5.0)
	1	T2-1	4900	5.88	Poisson Disk (d_{min}=4.3)
	2	T2-2	6400	7.68	Poisson Disk (d_{min}=3.8)
	3	T2-3	8100	9.72	Poisson Disk (d_{min}=3.3)
	4	T2-4	10000	12.00	Poisson Disk (d_{min}=3.0)
	5	T2-5	12100	14.52	Poisson Disk (d_{min}=2.7)
T3	0	T3-0	3600	4.32	Possion Distribution
	1	T3-1	4900	5.88	Possion Distribution
	2	T3-2	6400	7.68	Possion Distribution
	3	T3-3	8100	9.72	Possion Distribution
	4	T3-4	10000	12.00	Possion Distribution
	5	T3-5	12100	14.52	Possion Distribution
T4	0	T4-0	10000	12.00	Random Perturbation (d=1.0)
	1	T4-1	10000	12.00	Random Perturbation (d=1.5)
	2	T4-2	10000	12.00	Random Perturbation (d=2.0)
	3	T4-3	10000	12.00	Random Perturbation (d=2.5)
	4	T4-4	10000	12.00	Random Perturbation (d=3.0)
	5	T4-5	10000	12.00	Random Perturbation (d=3.5)

Table 4.3: Full description of the 4 simulated tissue sets used in the 4 experiments. Each set contains 6 types of tissues. A class index is an integer number assigned to the pixels of the classification map. The size of each simulated tissue field is $50mm \times 50mm$. The reflectivity distribution for all the simulated tissues is Gaussian with mean of 128 and standard deviation of 8.

density to moderately high scatterer density while the spatial randomness is kept to be fully random.

Tissue set T4 was used in experiment No. 4. In T4, the scatterer number density and scatterer reflectivity are kept fixed and the spatial randomness is controlled by random perturbation on regular lattice with mean perturbation vector length $\bar{d}$ changing from 0.1mm to 0.35mm. The purpose of this experiment is testing texture classification and segmentation of tissues deviating from complete anisotropic spatial regularity to certain degree of randomness.

4.7.2 Other Settings

For all the experiments, the imaging conditions are: Focused circular piston-like transducer with aperture radius 6.5mm, focal length 7.053cm, and central frequency 8Mhz; the transducer impulse response is simulated as a sine wave modulated by a Hanning window with bandwidth of 1Mhz; the excitation pulse is simulated as a sine wave modulated by a Hanning window with bandwidth of 4Mhz.

The performance of classification or segmentation is measured by *classification accuracy* which is defined as the percentage of pixels that have been assigned the correct class index. Since there is no other means to decide if a pixel has been correctly classified, the only criteria that can be used is the *a priori* known ground truth class map image that has been used to generate the tissue field. So the classification accuracy can be calculated by comparing the classified image pixels with the ground truth class map pixels. For each tissue set, a classification job was conducted to obtain the overall classification accuracy of all the 6 tissue types in that set. Several two-component texture based segmentation experiments were conducted to test the inter-class classification accuracy (among multi-classes) for all of the algorithms.

Let {B, L} denote the labels of tissue type for background (B) and lesion (L). For the two-component segmentation problem, the tissue combinations {B, L} chosen for background and lesion regions in a tissue set {0, 1, 2, 3, 4, 5} are: 3 pairs of adjacent tissue types ({B, L} = {0, 1}, {2, 3}, {4, 5}); 1 pair of tissue types with largest property difference on both ends of the tissue set "spectrum" ({B, L} = {0, 5}); 2 pairs of moderately separated tissue types ({B, L}= {0, 2}, {3, 5}).

4.8 Experimental Results

This section presents some important classification and segmentation results in various formats. The complete results data of this chapter can be found in Appendix E.

4.8.1 Classification and Segmentation Results In Tables

Tables 4.4, 4.5, 4.6, 4.7 summarize the classification and segmentation results of Experiments No. 1 to No. 4. The classification accuracy here is the overall (averaged) classification accuracy of all 6 tissue types in one tissue set.

Classification							
Tissue Set		ICA	PCA	Gabor	FWHM-ACVF	SNR-ALPHA	SNR-GAMMA
T1		82.3%	73.8%	65.6%	53.5%	47.2%	46.2%
Segmentation							
Tissue Set: T1		ICA	PCA	Gabor	FWHM-ACVF	SNR-ALPHA	SNR-GAMMA
Background	Lesion						
T1-0	T1-1	64.9%	56.8%	51.2%	53.5%	53.5%	48.9%
T1-2	T1-3	73.2%	63.5%	57.3%	24.1%	42.4%	45.1%
T1-4	T1-5	95.2%	86.9%	83.9%	64.6%	48.5%	50.1%
T1-0	T1-2	45.9%	47.7%	36.2%	50.0%	46.6%	44.7%
T1-3	T1-5	83.1%	78.8%	71.5%	60.8%	41.5%	42.3%
T1-0	T1-5	56.9%	57.0%	45.3%	51.8%	42.5%	40.4%

Table 4.4: Overall classification and two-component segmentation results of Exp. No. 1

Classification							
Tissue Set		ICA	PCA	Gabor	FWHM-ACVF	SNR-ALPHA	SNR-GAMMA
T2		45.7%	36.9%	31.5%	30.9%	51.8%	53.2%
Segmentation							
Tissue Set: T2		ICA	PCA	Gabor	FWHM-ACVF	SNR-ALPHA	SNR-GAMMA
Background	Lesion						
T2-0	T2-1	59.6%	51.4%	33.8%	42.8%	68.9%	70.0%
T2-2	T2-3	39.2%	24.5%	21.2%	5.6%	31.9%	36.0%
T2-4	T2-5	22.9%	23.6%	23.8%	37.6%	26.6%	30.6%
T2-0	T2-2	62.9%	55.0%	35.8%	32.0%	67.0%	68.2%
T2-3	T2-5	34.4%	22.4%	19.6%	34.5%	44.6%	42.2%
T2-0	T2-5	63.5%	55.0%	29.0%	27.3%	60.0%	60.2%

Table 4.5: Overall classification and two-component segmentation results of Exp. No. 2

Classification						
Tissue	ICA	PCA	Gabor	FWHM-ACVF	SNR-ALPHA	SNR-GAMMA
Tissue Set T3	26.8%	19.2%	19.4%	17.1%	44.3%	46.7%

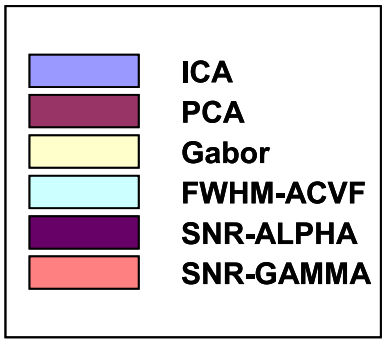
Table 4.6: Overall classification results of Exp. No. 3. The segmentation results are omitted here because the segmentation totally failed

Classification						
Tissue		ICA	PCA	Gabor	FWHM-ACVF	SNR-ALPHA	SNR-GAMMA
Tissue Set T4		78.5%	64.0%	74.3%	52.3%	43.5%	43.4%
Segmentation							
Tissue Set: T4		ICA	PCA	Gabor	FWHM-ACVF	SNR-ALPHA	SNR-GAMMA
Background	Lesion						
T4-0	T4-1	94.3%	85.0%	92.3%	18.7%	80.2%	78.4%
T4-2	T4-3	76.1%	68.8%	70.4%	59.9%	32.4%	32.0%
T4-4	T4-5	55.0%	56.2%	58.8%	39.4%	26.0%	12.9%
T4-0	T4-2	85.6%	80.7%	82.3%	21.2%	68.6%	68.7%
T4-3	T4-5	70.3%	57.5%	61.2%	58.8%	24.3%	28.0%
T4-0	T4-5	77.5%	72.0%	74.5%	15.4%	62.8%	62.6%

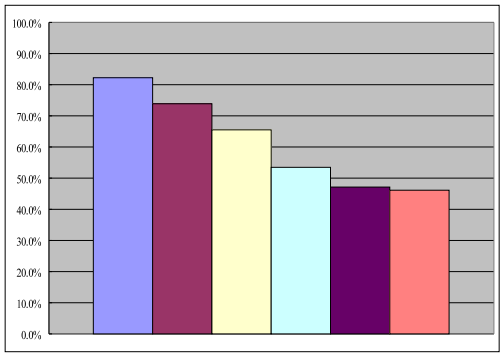
Table 4.7: Overall classification and two-component segmentation results of Exp. No. 4

4.8.2 Overall Classification Accuracy Comparison

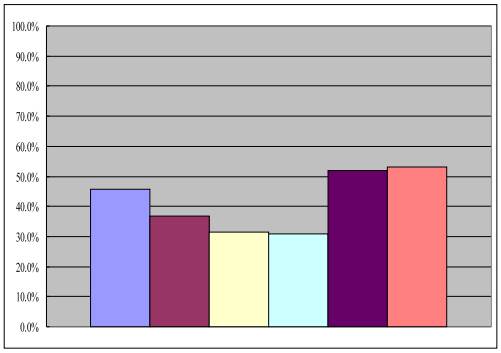
Figure 4.13 shows the bar charts of the overall classification accuracy using different algorithms.



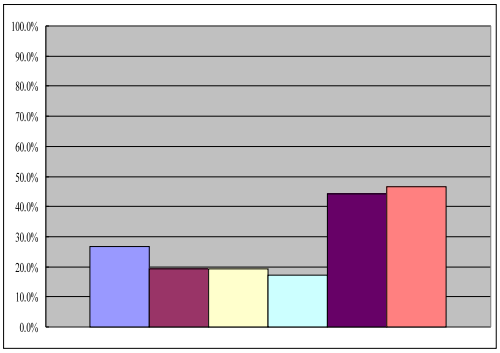
(a) Legend



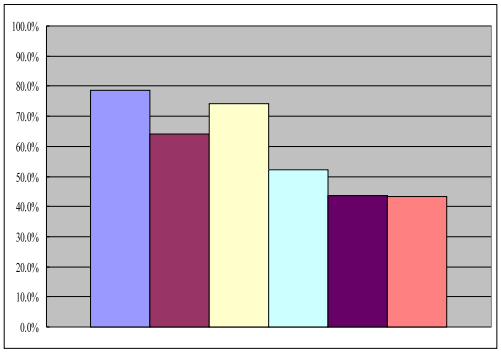
(b) Exp. No. 1



(c) Exp. No. 2



(d) Exp. No. 3

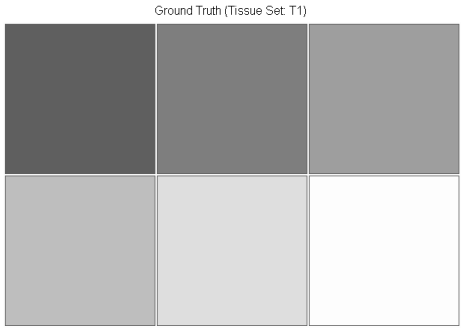


(e) Exp. No. 4

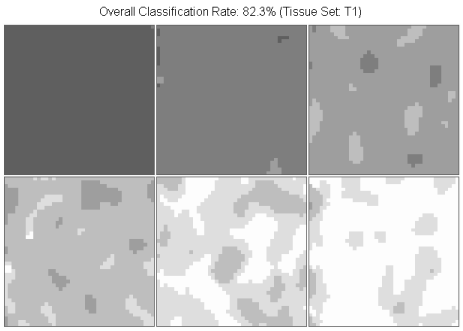
Figure 4.13: Overall classification results shown as bar chart.

4.8.3 An Example of Classification Map Using ICA

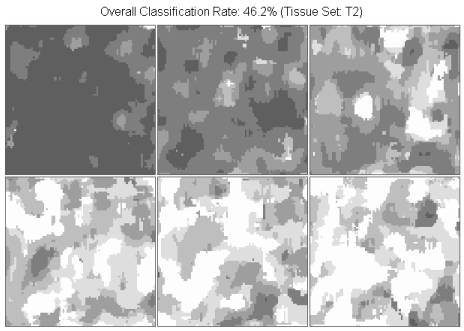
Figure 4.14 shows an example of the classification map for all 4 experiments using ICA. Sub-figure (a) shows the ground truth image of tissue types. Each grey-level of the image represents a tissue type. The darkest image corresponds to the non-Rayleigh region while the whitest image corresponds to the Rayleigh region. It can be observed that when the tissue type approaches Rayleigh region, the classification accuracy decreases.



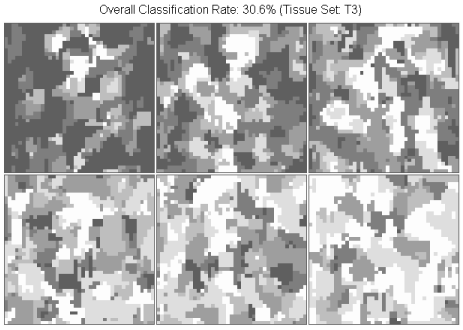
(a) Ground truth class map



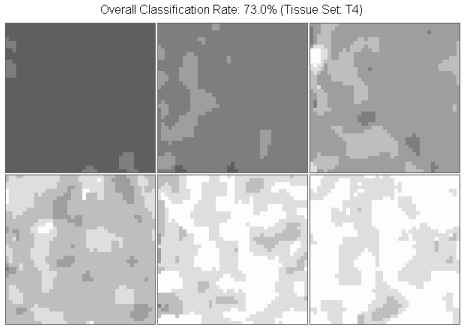
(b) Exp. No.1



(c) Exp. No.2



(d) Exp. No.3

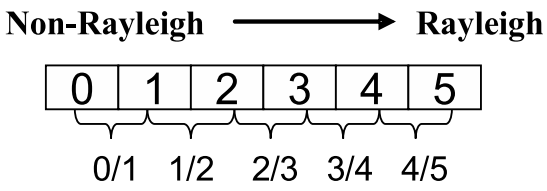


(e) Exp. No.4

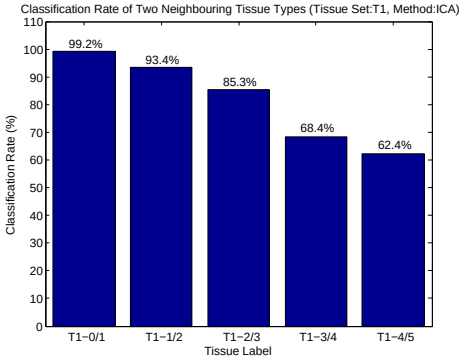
Figure 4.14: An example of classification map using ICA

4.8.4 Pairwise Tissue Classification Results Using ICA

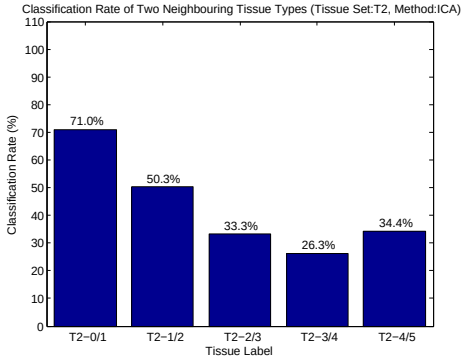
Figure 4.15 shows the pairwise classification results. Classification accuracy is computed for two neighboring tissue types which belong to the same tissue set. Different combinations of tissue pairs are given in (a) of the figure. One can clearly see that when the tissue type approaches Rayleigh limit, the classification accuracy drops quickly (One exception is Exp. No. 3).



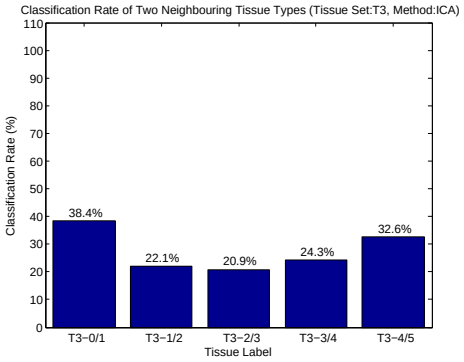
(a) Legend



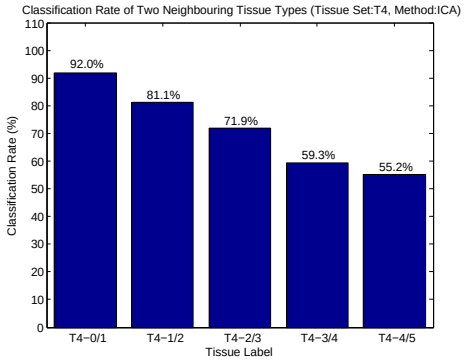
(b) Exp. No.1



(c) Exp. No.2



(d) Exp. No.3



(e) Exp. No.4

Figure 4.15: Pairwise tissue classification results using ICA. (a) shows the combinations of tissue pairs.

4.8.5 An Illustrative Example of Two-component Texture Segmentation Using Different Algorithms

Figure 4.16 shows an example of two-component texture segmentation using different algorithms. We start with a two-component template with a blob-shaped lesion region embedded in the background. In this example, the tissue type T1-0 was used for the lesion region and T1-2 was used for the background. A two-component speckle image is then generated based on the template. Two first order statistics algorithms, SNR-ALPHA and SNR-GAMMA, were applied to the image and segmentation results obtained. For the rest of four algorithms (ICA, PCA, Gabor and FWHM-ACVF), the original speckle image had to go through the preprocessing step before being processed by those algorithms. Actually those four algorithms do not rely on the first order statistical information of speckle so the preprocessing step makes it possible for us to compare the performance of those algorithms with the same starting point. If we apply SNR-ALPHA and SNR-GAMMA to the preprocessed image, the segmentation results are total failures since the first order statistical information (mean and variance here) has been lost.

4.8.6 Other Examples of Two-component Texture Segmentation

Figure 4.17 and 4.18 show two other examples of two-component texture segmentation. In Fig. 4.17, which the tissue types for background and lesion are adjacent ones, all the second and higher order statistics methods outperform first order statistics methods. In Fig. 4.18, the two first order statistics methods outperform second and higher order statistics.

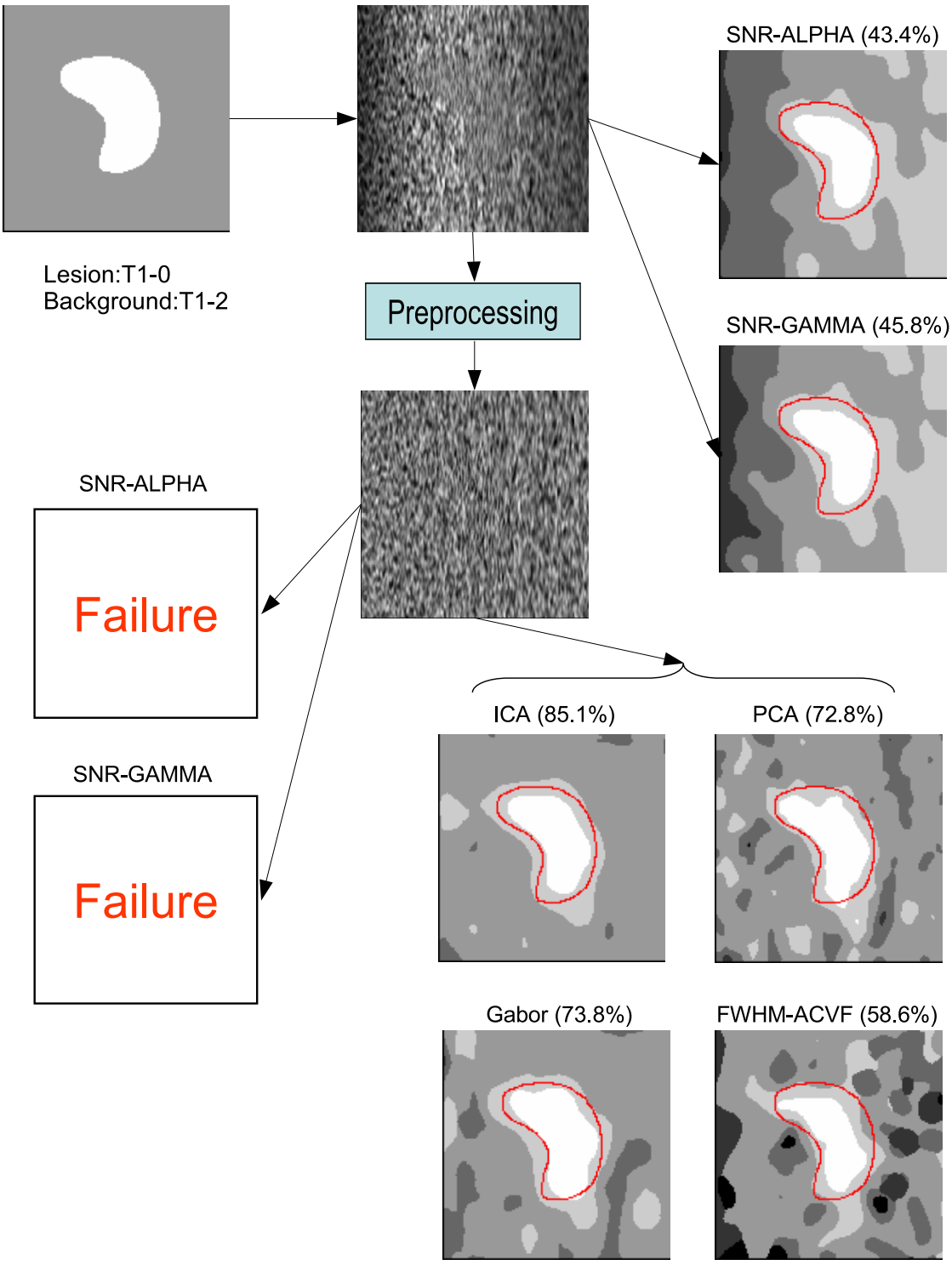
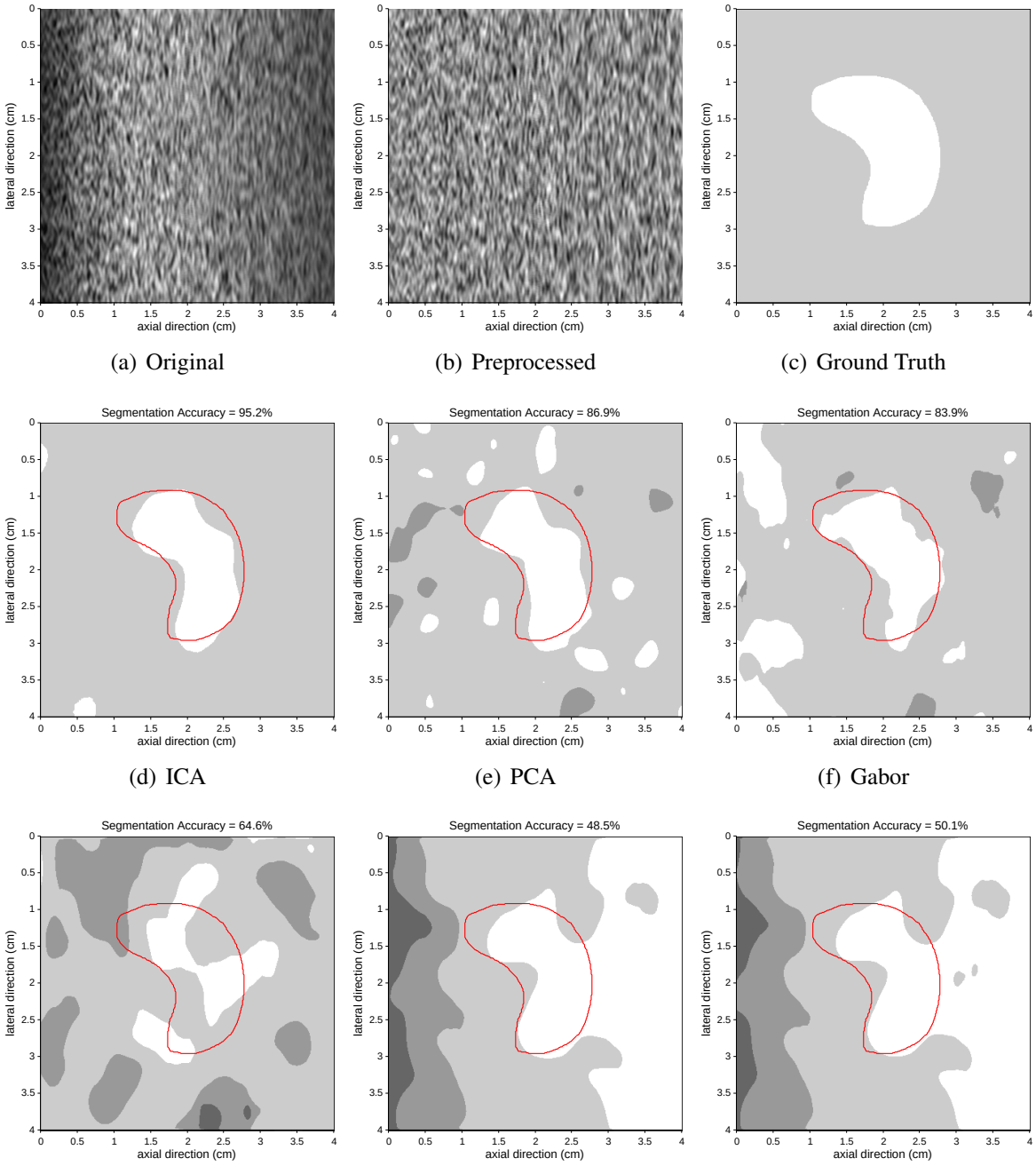


Figure 4.16: An example of two-component texture segmentation. This figure also shows the simplified work flow of different algorithms. For the two FOS algorithms SNR-ALPHA and SNR-GAMMA, we have to use the original speckle image without preprocessing. However the preprocessing step is necessary for the other algorithms (especially for ICA).

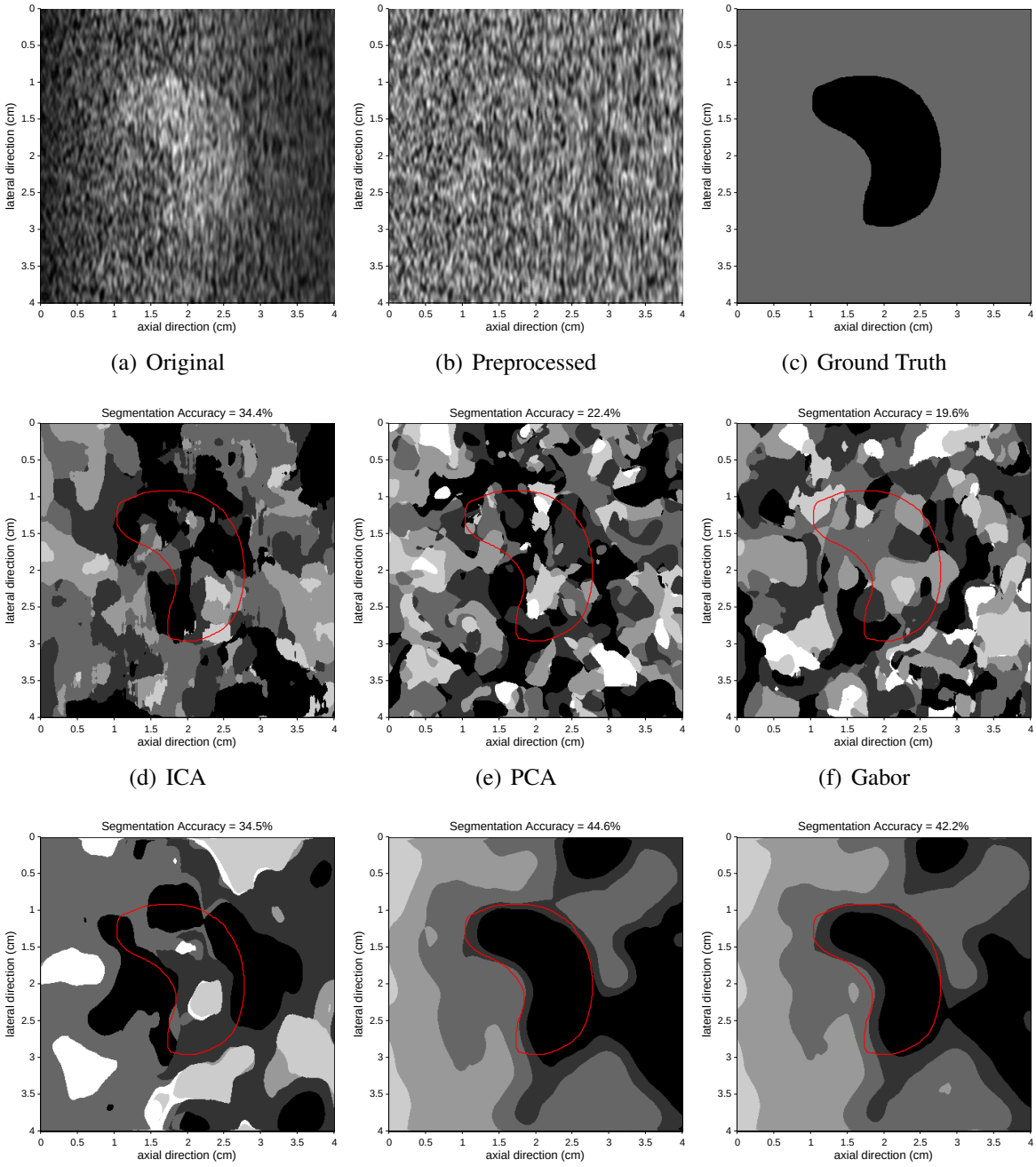


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM-ACVF (h) SNR-ALPHA (i) SNR-GAMMA

Figure 4.17: Two-component Texture Segmentation Results (Background:T1-4, Lesion:T1-5). (a) and (b) are the two-component speckle texture images before and after preprocessing. (d),(e),f and (g) are obtained from the (b); (h) and (i) are obtained from (a). This example demonstrated that when the background and lesion regions are very similar tissue types, second or higher statistics based methods outperform first order statistics based methods. In all the comparative second and higher order methods, ICA has shown to be the best performer.



(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure 4.18: Two-component Texture Segmentation Results (Background:T2-3, Lesion:T2-5) a) and (b) are the two-component speckle texture images before and after preprocessing. (d),(e),f and (g) are obtained from the (b); (h) and (i) are obtained from (a). This example demonstrated that in certain situations, the first order statistics may outperform second or higher order statistics based methods. In this example, the tissue types for background and lesion regions are not adjacent types and their first order statistics are quite different.

4.9 Discussions

In all four numerical experiments, ICA has shown to be a better performer than PCA, Gabor and FWHM-ACVF. Actually, in Exp. No. 1 and No. 4, ICA is the best performer in terms of overall classification accuracy. In Exp. No. 2 and No. 3, first order statistics algorithms SNR-ALPHA and SNR-GAMMA are best performers. The performances of PCA and Gabor approaches are consistently worse yet close to that of the ICA approach. The three multichannel filtering approaches, ICA, PCA and Gabor, are all better performers than the two first order statistics approaches in Exp. No. 1 and No. 4. This can be explained as that the second or higher order statistics may provide more information about the speckle texture than just the first order statistics. That ICA always outperforms PCA and FWHM-ACVF can be possibly explained as that HOS shows advantages over SOS. That ICA always outperforms Gabor can be explained as that the data-adaptive ICA bases have advantages over the fixed bases used in Gabor. Although both belong to second order statistics, the performance of FWHM-ACVF is consistently worse than that of PCA, which can be explained as that the former approach considers only the correlation on horizontal and vertical directions but the later one has no such restrictions.

In the segmentation experiments, the proposed ICA-based algorithm shows strong capability of segmenting textures without any mean and variance information (see Fig. 4.16 and 4.17). The boundary between the lesion and background edge is extremely fuzzy and it would be unrealistic to apply any conventional edge detection algorithm. For such type of problem, the only information that can be used is the texture. It seems that the proposed algorithm can nicely segment the lesion and background regions into different classes of tissues. Visually, the lesion object shape is better detected than any other algorithms. Although only two-component texture images are tested, the experiments conducted are actually multi-class texture segmentation which is more challenging than a conventional binary segmentation problem. In binary segmentation, the goal is to segment an image into two regions and it doesn't matter which class labels are assigned to them. In multi-class texture

segmentation, not only do the two or multiple lesion and background regions need to be detected but also the tissue classes they belong to should be decided. In the experiments, when the lesion and background belong to two non-adjacent tissue types, the fuzzy boundary of the lesion region has been classified as the tissue types in between. In fact, if the class label (or index) is replaced by some real valued physical measure of tissue, then the classification problem becomes a regression problem. Multi-dimensional regression could then be conducted to predict the tissue property (or properties) given as continuous real value(s).

Experiment No. 1 and No. 4 have been designed to test the situation when the spatial distribution of scatterers deviates from full randomness to certain degree of regularity (in both isotropic and anisotropic cases) while the scatterer number density is kept the same. It has been observed that the performance of Gabor in Exp. No. 1 is much worse than that of ICA (65.6% versus 82.3%), while in Exp. No. 4 the performance of Gabor is quite close to that of ICA (74.3% versus 78.5%). This can be explained as that the speckle texture in Exp. No. 4 is somewhat unnaturally regular than that in Exp. No. 1, so that the data adaptation nature of ICA does not show any obvious advantage over Gabor. In fact, some textures such as T4-0 and T4-1 (see Fig. E.41) look like artificially periodical textures instead of realistic ultrasound speckle texture.

Experiment No. 3 has been designed to test the situation when the spatial randomness is kept the same (homogenous Poisson process) and the scatterer number density deviates from very high density to moderately high density. This experiment shows that all of the three multichannel filtering methods and FWHM-ACVF failed to classify the tissues, while the two first order statistics methods become the best performers. Among the bad performers, ICA still performs slightly better than PCA and Gabor. FWHM-ACVF is the worst performer of all and its classification accuracy (17.1%) almost reaches the lower limit of a 6-class classification problem (16.6% = 100%/6). Since the tissue set T3 used in Exp. No. 3 are fully random scatterer fields with high scatterer density, the speckles are fully

developed. The possible explanation about the failure of ICA and other multichannel filtering algorithms is that when the speckle reaches the Rayleigh limit, the second or high order statistics also converges to a limit status which is independent of the scatterer number density. So that the classification capability of the second or high order statistics drops greatly. When the speckle reaches the Rayleigh limit, the only way to differentiate tissues may be the variance of the signal, which has been excluded from our experiments. The conclusion is that our experiments have demonstrated, although not analytically proved, that the second and high order statistics lost the texture classification capability when speckle reaches the Rayleigh limit. However, since the Rayleigh distribution is an extreme case and is simple to characterize, non-Rayleigh cases would be more of interest for real world problems because human soft tissues often have some microstructures with a certain degree of regularity. Consequently, the high order statistics based methods would still be useful.

The tissue set T2 used in Exp. No. 2 is similar to tissue set T3 for Exp. No. 3, but the difference is that in T2 the spatial randomness of scatterers (which is controlled by the Poisson disk process) is slightly more regular than T3. Experiment No. 2 has been designed to test the situation when the scatterer number density deviates from very high density to moderately high density while some level of regularity is kept for the scatterer spatial randomness. The results of Exp. No. 2 have shown that although the overall classification performance of the ICA-based approach is still worse than SNR-ALPHA and SNR-GAMMA, the relative performance improvement of the ICA-based approach is much higher than the two first order statistics methods when comparing with the results of Exp. No. 3 (for the ICA-approach, the performance has been improved from 26.8% to 45.7%, while for the SNR-ALPHA and SNR-GAMMA, the performance has been improved from about 45.0% to 52.0%). In some cases of two-component segmentation (see Fig. E.27 and Fig. E.30), the ICA-based approach even slightly outperforms SNR-ALPHA and SNR-GAMMA. For all the cases in Exp. No. 2, the ICA-based approach obviously outperforms PCA, Gabor and FWHM-ACVF. Based on above observations, one conclusion that can be drawn is that

if the spatial regularity slightly increases for speckles deviating from the Rayleigh model to non-Rayleigh model as the scatterer number density decreases, the performance of the ICA-based approach can be greatly improved.

The two first order statistics methods SNR-ALPHA and SNR-GAMMA show some stable classification capability for all the tissues tested. However, their classification rates are much lower than the ICA-based method in many cases except in some cases in Exp. No. 2 and No. 3 (See Fig. 4.18). These two methods don't work well for slightly non-stationary speckle texture images. Their classification map shows the band-shape artifacts along the lateral direction which is caused by the non-stationarity of an ultrasound image along the axial direction. An algorithm combining the proposed ICA-based approach with any proper first order statistics method could possibly be more adaptive to a wider range of real world problems and the classification accuracy is expected to be improved.

Chapter 5

ICA-based Ultrasound Image Despeckling

Speckle noise tends to reduce the image resolution and contrast, thereby reducing the diagnostic value of the ultrasound imaging. As a result, speckle reduction is always an important prerequisite for many other ultrasound image processing tasks. In this chapter, an ICA-based ultrasound despeckling approach is proposed by adapting the ICA Sparse Code Shrinkage (ICA-SCS) algorithm for the despeckling problem of ultrasound B-mode images with an appropriate preprocessing step suggested in [59]. ICS-SCS has been successfully used by the researchers for natural images denoising. The preprocessing step makes the ultrasound speckle noise much closer to real WGN hence more amenable to any denoising algorithms designed for WGN. Several classical methods have been chosen for comparison study purposes.

5.1 Ultrasound Image Formation Model

Assuming linear wave propagation and weak scattering, the back-scattered signal and the tissue reflectivity function are well-known to obey a simple Fourier transform relationship with respect to each other. In this case, a RF image is considered to be the convolution

of the point spread function (PSF) of imaging system with the tissue reflectivity function, which is given by

$$g(i, j) = f(i, j) \otimes h(i, j) + u(i, j) \tag{5.1}$$

where $g(i, j)$, $f(i, j)$, $h(i, j)$ and $u(i, j)$ denote the RF image, the tissue reflectivity function, the PSF and the additive noise not accounted for by the convolution model. i and j denote the axial and lateral indices of the image samples. The corresponding power spectral density of the RF image is given by

$$P_g(\omega_1, \omega_2) = P_f(\omega_1, \omega_2)|H(\omega_1, \omega_2)|^2 + P_u(\omega_1, \omega_2) \tag{5.2}$$

Usually the samples of tissue reflectivity function $f(i, j)$ and the additive noise $u(i, j)$ can be assumed to be uncorrelated. Then the power spectral densities of them become constant σ_f^2 and σ_u^2, which are the variances of the reflectivity function and additive noise respectively. Now Eq. 5.2 can be rewritten as

$$P_g(\omega_1, \omega_2) = \sigma_f^2|H(\omega_1, \omega_2)|^2 + \sigma_u^2 \tag{5.3}$$

When a RF image is envelope detected using the Hilbert transform (See Appendix D), the frequency demodulated in-phase/quadrature (I/Q) image can still be modeled as a convolution mixture of a complex reflectivity function with a complex PSF due to the linearity of the frequency demodulation process. Hence the Eq. 5.1 can still be used.

5.2 Ultrasound Speckle Noise Model

The speckle noise in ultrasound envelope images (before logarithmic transformation), can be modeled as multiplicative in nature [81], given by

$$y(i, j) = x(i, j)n(i, j) + a(i, j) \tag{5.4}$$

where $y(i, j)$ represents the noise pixel in the middle of a moving window, $x(i, j)$ represents the corresponding noise-free pixel, $n(i, j)$ and $a(i, j)$ represents the multiplicative and additive noise, respectively, and i, j are the indices of the spatial locations. If logarithmic transformation is applied to the image, the speckle statistics will be changed. One change is that the local mean becomes proportional to the local variance rather than the standard deviation. More specifically, logarithmic transformation affects the high-intensity tail of the Rayleigh, K and Rice *pdf*s more than it does the low-intensity part. As a result, the speckle noise becomes closer to Gaussian noise [15]. Because the effect of additive noise is usually much smaller compared with that of multiplicative noise, Eq. 5.4 can be written as

$$y(i, j) \approx x(i, j)n(i, j) \tag{5.5}$$

After taking logarithmic transformation, the multiplicative noise can then be converted into additive noise form as

$$\log(y(i, j)) = \log(x(i, j)) + \log(n(i, j)) \tag{5.6}$$

Among many methods that have been proposed to perform the despeckling task, homomorphic despeckling methods are a class of approaches that take advantage of the model given in Eq. 5.6. The common assumption made in a dominant number of such studies is that the samples of the additive noise are mutually uncorrelated and obey Gaussian distribution. Some conventional image restoration methods, such as Wiener filtering, wavelet filtering and anisotropic diffusion, can then be applied. The ICA-based despeckling method that will be investigated in this chapter also relies on this model.

5.3 Preprocessing Algorithm

Although Eq. 5.6 is frequently used, many current studies show conceptually and experimentally that this assumption is oversimplified and unnatural. Moreover, it may lead to inadequate performance of the speckle reduction methods. In the most typical practical setting, the speckle noise is neither Gaussian nor white, and it is spiky in nature. If not properly treated, the spiky component of such a noise can be processed by a denoising algorithm as a part of the useful signal that needs to be recovered. It allows a significant portion of the noise to be preserved by denoising, thereby considerably decreasing the efficiency. Before applying the ICA-based despeckling algorithm introduced in the next section, a preprocessing algorithm proposed in [59] needs to be applied to modify the speckle images (without affecting the anatomical information they contain) so that the speckle noise in the log-transformation domain becomes very close to real WGN. Then the ICA despeckling algorithm can perform under nearly optimal conditions. The preprocessing algorithm consists of two steps:

Step 1: An envelope detected image is subject to a spectrum equalization procedure intended to decorrelate the image samples. A linear filter $l(i, j)$ is defined to "flatten" the power spectral density of the image. The transfer function of the filter is given by:

$$L(\omega_1, \omega_2) = (|H(\omega_1, \omega_2)|^2 + \varepsilon)^{-1/2} \tag{5.7}$$

where $L(\omega_1, \omega_2)$ and $H(\omega_1, \omega_2)$ are the Fourier spectrum of the linear filter and the complex point spread function (PSF) of the envelope detected image. The constant ε is a tunable parameter to control the decorrelation extent. The magnitude of the spectrum $H(\omega_1, \omega_2)$ can be estimated by using the following described method. If the additive noise is ignored in Eq. 5.1, the log-magnitude of the Fourier transform of the envelope detected image can

be expressed as

$$G_l(\omega_1, \omega_2) = H_l(\omega_1, \omega_2) + F_l(\omega_1, \omega_2) \tag{5.8}$$

where $G_l(\omega_1, \omega_2)$, $H_l(\omega_1, \omega_2)$ and $F_l(\omega_1, \omega_2)$ are the log-magnitude of the Fourier transform of the envelope image, the complex tissue reflectivity function and the complex PSF respectively. This equation suggested that $|H(\omega_1, \omega_2)|$ in Eq. 5.7 can be estimated from $G_l(\omega_1, \omega_2)$ by rejecting the "noise" term $F_l(\omega_1, \omega_2)$. Thus the problem of estimating the power spectrum of the PSF becomes a filtering problem. If the reflectivity function behaves as WGN, then the samples of $F_l(\omega_1, \omega_2)$ are *i.i.d* and obey the Fisher-Tippett distribution, whose *pdf* is given by

$$p(x) = 2\exp\{(2x - \ln 2\sigma_f^2) - \exp\{2x - \ln 2\sigma_f^2\}\} \tag{5.9}$$

where σ_f^2 denotes the variance of the reflectivity function. The Fisher-Tippett distribution is close to Gaussian but has a long heavy tail. So it can be regarded as a WGN contaminated with outliers. Figure 5.1 shows the Fisher-Tippett *pdf* with σ_f^2 equals to 1. In order to reject

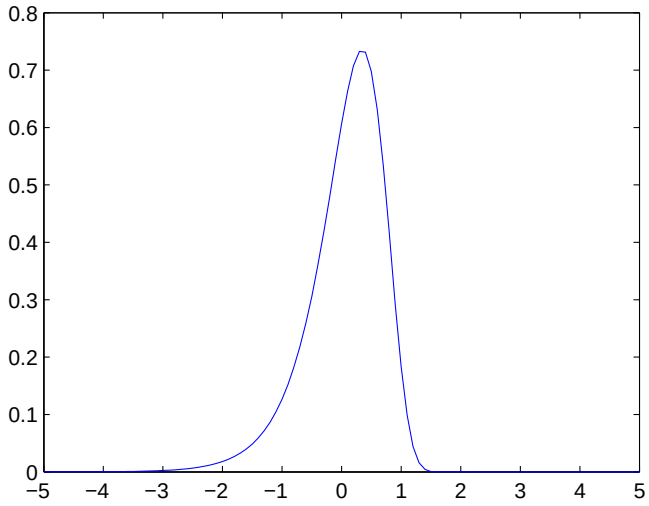


Figure 5.1: Fisher-Tippett *pdf* with $\sigma_f^2 = 1$

the spiky outliers in the Fish-Tippett distribution, $G_l(\omega_1, \omega_2)$ is passed through a nonlinear outlier-shrinkage procedure, whose purpose is to suppress the spiky component of the log-

transformed speckle noise. The spiky component can be estimated as robust residuals of $G_l(\omega_1, \omega_2)$ as

$$R(\omega_1, \omega_2) = \text{sign}(\Delta G_l(\omega_1, \omega_2))(\Delta G_l(\omega_1, \omega_2) - \lambda)_+ \tag{5.10}$$

where $\Delta G_l(\omega_1, \omega_2)$ denotes the difference between $G_l(\omega_1, \omega_2)$ and its median filtered version. λ is a pre-defined threshold, and the operator $(x)_+$ returns x if $x > 0$ and zero otherwise. By subtracting the spiky component from G_l, it can be demonstrated that the noise in the difference image $G_l - R$ behaves very similarly to real WGN [58]. The difference image $G_l - R$ can then be denoised using any classical method such as wavelet transform. The denoised image is actually an estimate of $H_l(\omega_1, \omega_2)$ and then the magnitude of PSF power spectrum $|H(\omega_1, \omega_2)|^2$ can be obtained correspondingly. Then the linear filter defined in Eq. 5.7 can be applied to the envelope detected image for decorrelation.

Step 2: After the envelope detected image has been decorrelated, the logarithmic transformation is conducted to convert the multiplicative speckle noise into additive noise. At this stage the noise has already been decorrelated (whitened). Usually the *pdf* shape of logarithmic transformed speckle resembles that of the Fisher-Tippett *pdf*. So the speckle noise is similar to WGN contaminated by a small number of outliers. The same outlier shrinkage method used in Step 1 can be used again to reduce the outliers so as to "Gaussianize" the speckle noise.

In later sections, the short name "W&G" ("Whitening and Gaussianization") will be given to this preprocessing algorithm. Figure 5.2 and 5.3 show the log-transformed speckle images before and after the preprocessing step. Figure 5.4 and 5.5 show the corresponding image autocorrelation function slices before and after preprocessing for axial and lateral directions respectively. Figure 5.6 shows the histogram of the log-transformed images before and after the preprocessing. It can be observed that the autocorrelation function slices on both vertical and horizontal become wider, indicating that the image has been to some

extent decorrelated. The histogram before the preprocessing is more like the Fisher-Tippett distribution while the histogram after preprocessing is more like a Gaussian.

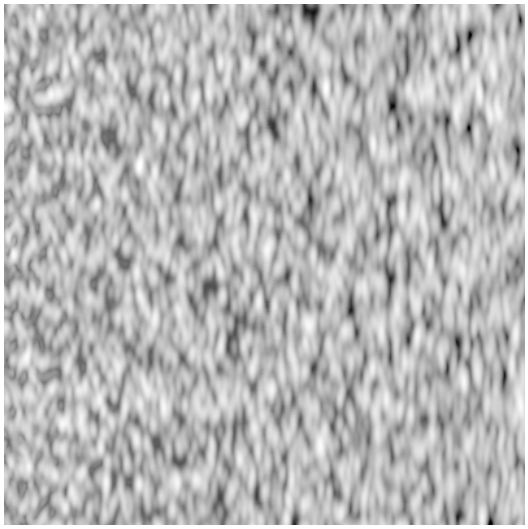


Figure 5.2: Speckle image before preprocessing.

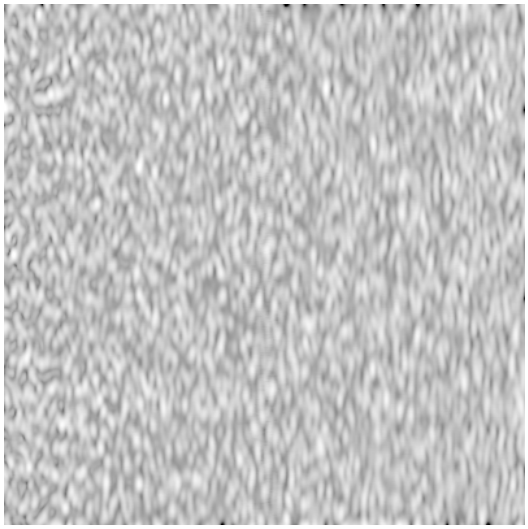


Figure 5.3: Speckle image after preprocessing.

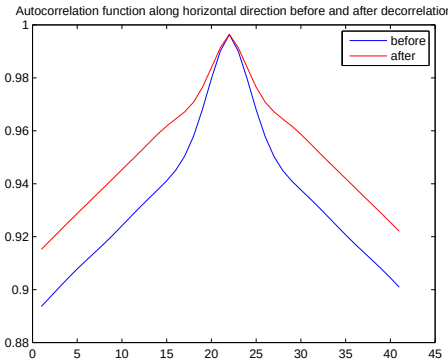


Figure 5.4: Axial slice of image autocorrelation function before and after preprocessing.

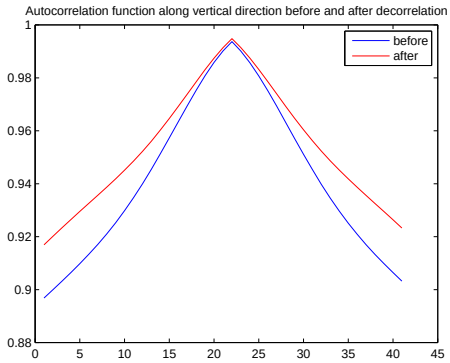


Figure 5.5: Lateral slice of image autocorrelation function before and after preprocessing.

5.4 Despeckling by ICA Sparse Code Shrinkage (ICA-SCS)

ICA has been used by Hyvärinen *et al.* [27] [29] to denoise non-Gaussian signal contaminated by additive WGN and has proved to be a promising method which outperforms many

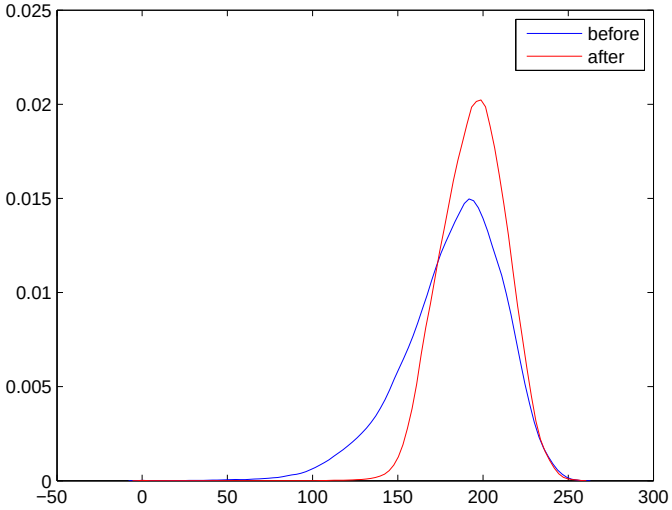


Figure 5.6: Histogram of the log-transformed speckle image before and after preprocessing.

traditional methods. Considering the ultrasound image speckle model in Eq. 5.6, this ICA based approach could possibly be used for the ultrasound image despeckling problem. The ICA spare code shrinkage method is closely related to wavelet soft thresholding, which has been used by many researchers for ultrasound image despeckling. Compared with wavelet transform, this ICA-based method has an important advantage since the features are fully adapted to the training images and determined solely by their statistical properties, while the wavelet transformation relies heavily on certain abstract mathematical properties that may be only weakly related to the properties of the image. A brief review of the theory of ICA spare code shrinkage is given in this section.

5.4.1 Scalar Random Variable Denoising

If an observed scalar random variable x can be modeled as

$$x = s + n \tag{5.11}$$

where s is the signal and n is noise, independent of s. Then the denoising problem is to find an estimate

$$\hat{s} = g(x) \tag{5.12}$$

using only the statistical properties of s and n, such that $\hat{s}$ is as close to s as possible, in some well-defined sense.

The denoising problem can be formed as choosing an estimator for a random variable with a known probability density. *Minimum Mean Squared Error* (MMSE) estimator and *Maximum a posteriori* (MAP) estimators are two of the most classical estimators. A typical MAP estimator can be written as

$$\hat{s} = \arg\max_{s} p(s|x) \tag{5.13}$$

If the signal has sparse (super-Gaussian) density function, it can be modeled parametrically by using functions such as *Laplacian* (also called double-sided exponential), which is given by

$$p(s) = \frac{1}{\sqrt{2}d} \exp(-\sqrt{2}|s|/d) \tag{5.14}$$

This function has only one parameter, the variance d^2. It can be combined with Gaussian to represent broader varieties of density functions in this form

$$p(s) = C \exp(-as^2/2 - b|s|) \tag{5.15}$$

where C is a normalization constant, and a and b can be estimated as

$$b = \frac{2p_s(0)\mathrm{E}\{s^2\} - \mathrm{E}\{|s|\}}{\mathrm{E}\{s^2\} - (\mathrm{E}\{|s|\})^2} \tag{5.16}$$

$$a = \frac{1}{\mathrm{E}\{s^2\}}(1 - E\{|s|\}b) \tag{5.17}$$

Based on the density model in Eq. 5.15 and the MAP estimator in Eq. 5.13, assuming additive Gaussian noise: $n \sim N(0, \sigma^2)$, the closed form of the estimator $\hat{s}$ can be derived, using Bayes rule, as below

$$\hat{s} = g(x) = \frac{1}{1 + \sigma^2 a} \text{sign}(x) \max(0, |x| - b\sigma^2) \tag{5.18}$$

This function is called a *shrinkage* function. When $a = 0$ (assuming Laplacian variable) then this function is the same form as the *soft thresholding* function introduced by Donoho [14] for wavelet denoising. Figure 5.7 shows the sparse density function defined in Eq. 5.15 and Fig. 5.8 shows the corresponding shrinkage function defined in Eq. 5.18 for $a = b = 0.5$ and $\sigma = 1$..

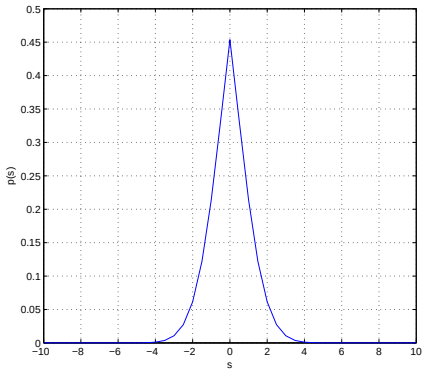


Figure 5.7: Sparse density function combined Laplacian and Gaussian defined in Eq. 5.15.

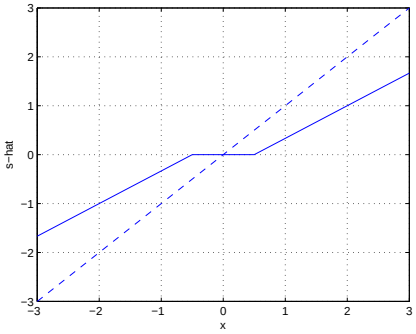


Figure 5.8: Shrinkage function defined in Eq. 5.18, with $a = b = 0.5$ and $\sigma = 1$. Function $\hat{s} = x$ is drawn for comparison.

For very sparse density, Hyvärinen [29] introduced the following model

$$p(s) = \frac{1}{2d} \frac{(\alpha + 2)[\alpha(\alpha + 1)/2]^{(\alpha/2+1)}}{[\sqrt{\alpha(\alpha + 1)/2} + |s/d|]^{(\alpha+3)}} \tag{5.19}$$

where d is the standard deviation and α is a parameter which controls the sparseness. When $\alpha \to \infty$, the limit is Laplace density. Since d is directly the standard deviation of the model,

it is logical to set d to the estimated standard deviation of the data. α can be derived as

$$\alpha = \frac{2 - k + \sqrt{k(k+4)}}{2k-1} \tag{5.20}$$

where $k = d^2 p(0)^2$. Thus, one can estimate α by first estimating $p(0)$ and d (In our numerical experiment, the MATLAB kernel-based density estimation function *ksdensity* is used for estimating $p(0)$). The MAP estimator (shrinkage function) of this model can be found as

$$\hat{s} = g(x) = \text{sign}(x) \max[0, \frac{|x| - ad}{2} + \frac{1}{2}\sqrt{(|x| + ad)^2 - 4\sigma^2(\alpha+3)}] \tag{5.21}$$

where $a = \sqrt{\alpha(\alpha+1)/2}$ and $g(x)$ is set to zero whenever the square root in Eq. 5.21 is imaginary.

5.4.2 Random Vector Denoising

Having discussed the denoising of a scalar variable, now let us consider an observed random vector $\mathbf{x}$ which can be modeled as

$$\mathbf{x} = \mathbf{v} + \mathbf{n} \tag{5.22}$$

where random vector $\mathbf{n}$ is white Gaussian noise independent of signal random vector $\mathbf{v}$. The goal is to find $\hat{\mathbf{v}} = \mathbf{g}(\mathbf{x})$ such that $\hat{\mathbf{v}}$ is close to $\mathbf{v}$ in some well defined sense. Theoretically we can define the MMSE and MAP estimators in a similar way as we define them for a scalar variable. Since usually there are complex statistical dependencies among the components, denoising based on individual components is a very crude approximation. If we try to define the estimator in high dimension, again we encounter the "curse of dimensionality" of estimating high dimensional density.

If a properly chosen linear transform is applied to $\mathbf{x}$ to decompose it into sparse and

independent components as below

$$\mathbf{Wx} = \mathbf{Wv} + \mathbf{Wn} \tag{5.23}$$

then the scalar variable denoising method can be used for each individual component of the transformed vector. For any $\mathbf{W}$, $\mathbf{Wn}$ remains to be Gaussian and if $\mathbf{W}$ is also orthogonal then $\mathbf{Wn}$ remains to be white. ICA transform, after orthogonalization, is exactly such a linear transform which can decompose the random vector signal into independent components with sparsity while still maintaining the statistical properties of the noise. Then it is possible to filter out the noise through shrinkage of each individual component.

5.4.3 Summarization of ICA-SCS Algorithm

The ICA-SCS denoising algorithm can be summarized as

1. Estimate the ICA transform matrix $\mathbf{W}$ from noise free data vector $\mathbf{z}$ with the same statistical properties as the data that we wish to denoise. Then $\mathbf{W}$ is orthogonalized as $\mathbf{W} \leftarrow \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1/2}$.

2. For every component $s_i = \mathbf{w}_i^T\mathbf{z}$, estimate a parametric density model and the corresponding shrinkage nonlinearity function g_i.

3. For observed noisy vector $\mathbf{x}$, transform it using $\mathbf{W}$. Apply the shrinkage function g_i obtained in step 2 to each component of the transformed vector to get $\hat{s}_i = g_i([\mathbf{Wx}]_i)$.

4. Reconstruct the denoised vector by inverse transform: $\hat{\mathbf{v}} = \mathbf{W}^{-1}\hat{\mathbf{s}}$.

5.4.4 Obtaining Speckle-free Training Images

ICA-SCS algorithm has been successfully applied to the denoising problem of natural images with the noise assumed to be additive WGN. Since the preprocessing step described in section 5.3 has been used to make the speckle much closer to real WGN, the performance of ICA-SCS should be greatly improved. But one shall notice that this algorithm requires obtaining some noise-free training images to estimate the ICA transform and nonlinearity shrinkage function. In ultrasound imaging it is not easy to obtain absolute noise-free (speckle-free) training images so it seems that the algorithm can not be used directly. We need either to find a proper model for the noise-free images or to investigate any modified approaches that do not require using noise-free images for training. However, in practice ICA-SCS still can be used if any existing techniques can be used to preprocess the noisy training images and obtain the approximate parameters for shrinkage function. In our numerical study, the *a priori* known tissue reflectivity functions used for simulating the B-mode images are used as noise-free images for convenience. In reality, the noise-free image (or image patches) could possibly be obtained through the following methods:

1. Obtaining speckle-free training images by using classical filtering techniques such as wavelet filtering, Wiener filtering, etc., to preprocess the speckle images. This method is easy to realize but the disadvantage is that many of the existing filtering techniques can not guarantee the best results. This may affect the estimation of the density model and the shrinkage nonlinearity function.

2. Obtaining noise-free training images by using frequency compounding or spatial compounding method involving multiple speckle images of the same tissue. Compounding method has been used as a state-of-the-art approach in industry and has very good speckle reduction performance. However, the disadvantage of this method is that multiple images are needed so the imaging process becomes complicated. The registration of multiple im-

ages is another difficult issue that needs to be tackled.

3. Obtaining the noise-free image samples (patches), instead of a whole noise-free image, by letting the experienced clinical radiologists delineate the anatomical details in the speckle images. Then through some simple interactive image processing, the despeckled image patches can be obtained from the delineated images. This is an empirical way to find the speckle-free image samples.

4. The tissue images obtained from other imaging modalities such as X-ray and MRI images can possibly be used as references to compare with the ultrasound images for obtaining speckle-free image patches. This topic needs to be investigated in future research.

For a specific despeckling task, training image patches can be collected accumulatively and saved in a database. More training samples can be added to the database and some unwanted samples can be removed from it. Over a long time of iterative processes, a good and stable collection of speckle-free image samples may be obtained.

5.5 Proposed ICA-based Ultrasound Despeckling Algorithm

The proposed algorithm is a combination of the W&G preprocessing algorithm and the ICA-SCS algorithm introduced previously. In order to avoid the unstability of an ICA algorithm when dealing with slightly non-stationary ultrasound images, the same histogram regulation step introduced in Chapter 4 is also used here. The diagram of the entire algorithm is shown in Fig. 5.9.

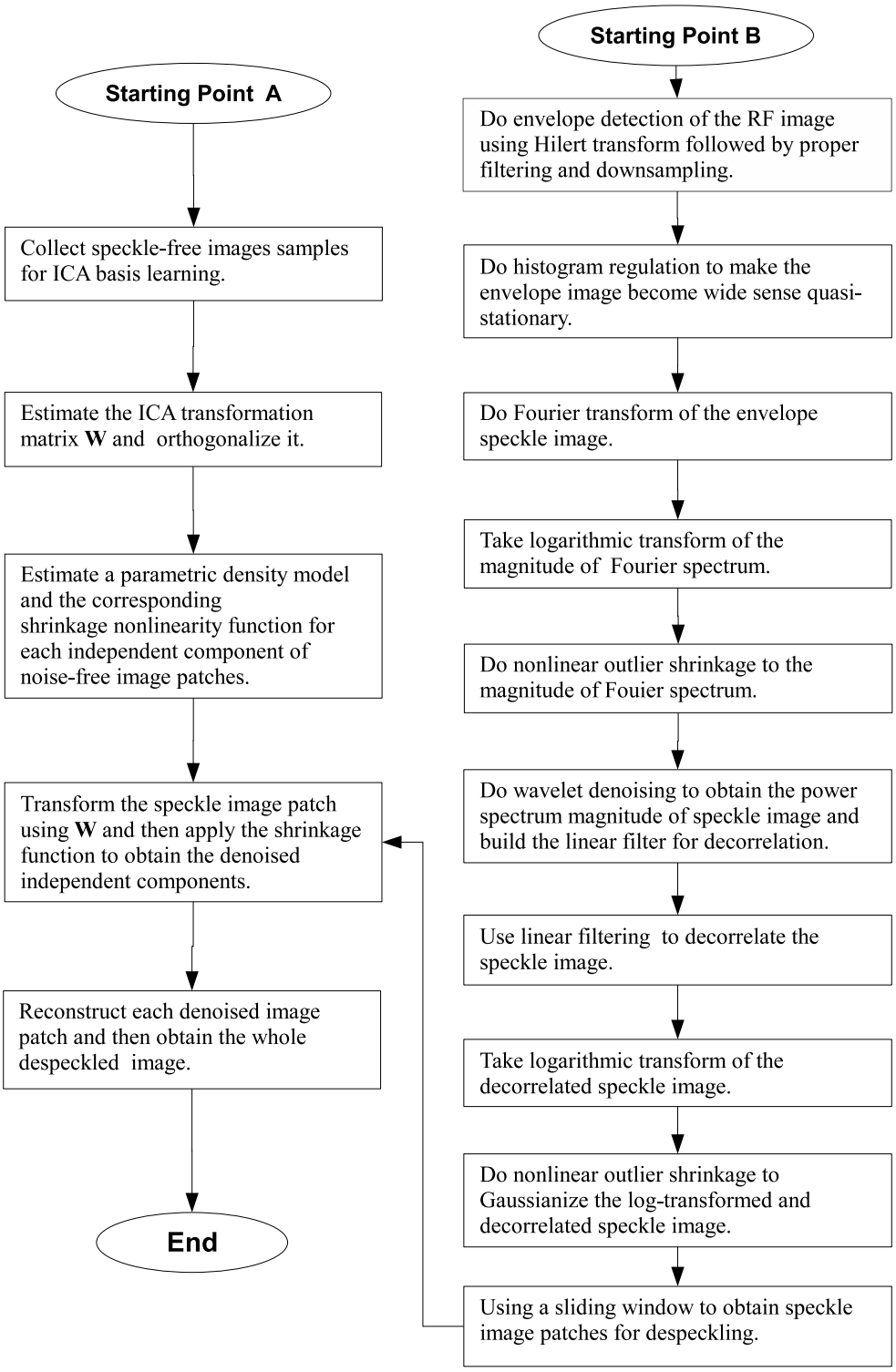


Figure 5.9: Diagram of the proposed ICA-based ultrasound despeckling algorithm.

5.6 Algorithms Chosen for Comparison Study

Several classical denoising algorithms have been chosen for comparison study. Those algorithms include both the classical filtering methods based on low order statistics as well as the sophisticated multichannel filtering method.

1. *2-D adaptive Wiener filtering*. 2-D adaptive Wiener filtering lowpass-filters a gray scale image that has been degraded by constant power additive noise. It uses a pixelwise adaptive Wiener method based on statistics estimated from a local neighborhood of each pixel. The local mean and variance can be estimated as

$$\mu = \frac{1}{n_x n_y} \sum_{i,j \in \mathrm{N}} x(i, j) \tag{5.24}$$

$$\sigma^2 = \frac{1}{n_x n_y} \sum_{i,j \in \mathrm{N}} [(x(i, j))^2 - \mu^2] \tag{5.25}$$

where N is an image window with size n_x and n_y, $x(i, j)$ are the pixels in the window. A pixelwise Wiener filter is then created using the estimates above

$$\hat{x}(i, j) = \mu + \frac{\sigma^2 - \nu^2}{\sigma^2}(x(i, j) - \mu) \tag{5.26}$$

where $\hat{x}(i, j)$ is the filtered pixel and ν^2 is the noise variance which is either given as a fixed value or can be estimated as the average of all local estimated variances.

2. *Wavelet shrinkage denoising by soft-thresholding*. This is a classical multichannel filtering method for WGN denoising, which is originally proposed by Donoho [14]. The basic idea of this method is transforming the noise signal to a wavelet decomposition and performing a shrinkage nonlinearity componentwise, then taking the inverse wavelet transform to obtain the denoised signal. The nonlinearity g originally proposed was a soft threshold-

ing function

$$g(x) = \text{sign}(x)\max(0, |x| - t) \tag{5.27}$$

where t is the shrinkage amount. Donoho suggested the universal threshold value $t = \sqrt{2\ln n} \cdot \sigma$ where n is the length of the data and σ is the standard deviation of the noise that can be estimated from the data. t is the optimal threshold in the asymptotic sense and minimizes the cost function of the difference between the function and the soft thresholded version of the same in the L2 norm sense. Compared to ICA-SCS, basic wavelet shrinkage uses a fixed transform whereas ICA-SCS uses an adaptive transform which is estimated from the data. The basic wavelet shrinkage algorithm has a fixed nonlinearity, whereas in ICA-SCS the nonlinearity is tailored to the distributions involved.

3. *L_2-mean filtering*. L_2-mean filtering is a classical despeckling technique commonly used in the ultrasound community [79]. For an image window N = (n_x, n_y), with n_x and n_y to be odd integers, the filter is defined as

$$\hat{x}(i_c, j_c) = \sqrt{\frac{1}{n_x n_y}\sum_{i,j\in \mathrm{N}}[x(i,j)]^2} \tag{5.28}$$

where $x(n, m)$ is the image pixel inside the window N with i and j as the image coordinates index. $\hat{x}(i_c, j_c)$ is the filtered center pixel of the window N, with i_c and j_c as the center coordinates indices. It has been shown that the L_2-mean filter is equivalent to the maximum likelihood estimation in the case that the speckle has (multiplicative) Rayleigh *pdf* [45].

5.7 Numerical Experiments and Results

Several numerical experiments were designed to test the despeckling of images containing some thin lines. The training images for ICA filter learning were generated using the procedure introduced in Chapter 3. The training images are speckle-free and contain some

thin straight lines of different skewing angles ranging from 0° to 180° in the step of 22.5°. Fig. 5.10 shows the learned ICA basis functions. The simulated speckle images for testing

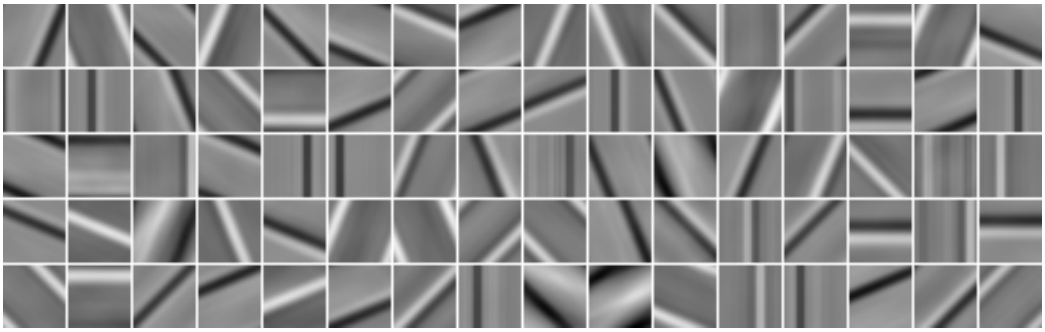


Figure 5.10: ICA basis functions learned from noise-free training images.

contain some naturally shaped thin lines with approximately the same width as those in the training images. These thin lines are used to represent some anatomical structures such as blood vessels found in real ultrasound images. Four experiments have been designed to test the performance of algorithms for speckle images with combinations of two noise levels and two speckle types. The ratios of the mean value of the scatterer reflectivity functions used for the signal (thin lines) and the noise (speckle) are set to 1.0 for Exp. No. 1 and No. 3, and 0.5 for Exp. No. 2 and No. 4, so that the noise level in Exp. No. 2 and No. 4 becomes about 2 times that of the noise level in Exp. No. 1 and No. 3. The signal level are kept the same for all four experiments. In Exp. No. 1 and No. 2, the speckles are fully developed while in Exp. No. 3 and No. 4, the speckles contain some coherent components which were simulated via the Poisson disk distribution.

In order to quantitatively compare the performance of different algorithms, several performance measures are defined. The first one is the standard speckle SNR which is defined as the ratio of mean to standard deviation for an image window. For fully developed speckle, this ratio should be close to 1.91 (See Eq. 3.6). When speckle is reduced, this value is expected to increase. The second measure is the normalized mean-square-error

(NMSE) which is defined as:

$$NMSE = \frac{\sqrt{\sum_{i,j}[I_0(i,j) - I(i,j)]^2}}{\sqrt{\sum_{i,j}[I_0(i,j)]^2}} \tag{5.29}$$

where I_0 and I are image windows in original ground truth image and the despeckled image, i, j are the coordinates index in the window. This performance measure is commonly used in the coherent imaging since the standard definition of signal-to-noise ratio might be inadequate for speckle with a multiplicative nature. The smaller this value is, the more the speckles are reduced.

The two measures above are intended to evaluate the degree of speckle reduction only and they do not consider how to evaluate the ability of sharp edge preservation. To solve this issue, the third measure ρ [19] can be defined as:

$$\rho = \frac{\sum_{i,j}[J_0(i,j) - \bar{J}_0][J(i,j) - \bar{J}]}{\sqrt{\sum_{i,j}[J_0(i,j) - \bar{J}_0]^2}\sqrt{\sum_{i,j}[J(i,j) - \bar{J}]^2}} \tag{5.30}$$

where J_0 and J are the high-pass filtered versions of the image window I_0 in original ground truth image and I in despeckled image, respectively. $\bar{J}_0$ and $\bar{J}$ are the corresponding mean values of J_0 and J. J_0 and J are obtained by applying a standard 3×3 Laplacian filter to I_0 and I. Image edges are better preserved when ρ is closing to 1.

In order to compute the performance values using the three measures defined above, multiple image windows at random locations are selected and the computed performance values are averaged out for all windows. Each algorithm for comparison has been tuned to its best performance status so that the speckle can be effectively reduced without causing obvious oversmoothing of the signal components (thin lines). For the wavelet shrinkage algorithm, 3-level wavelet decomposition based on the nearly symmetric wavelet with six vanishing moments was used. The performance measures are listed in Tables 5.1 to 5.4. Many corresponding images are listed in Figures 5.11 to 5.19.

With W&G Preprocessing					
	ICA-SCS	Wavelet Shrinkage	Adaptive 2-D Wiener	L2-mean filter	Speckle Image
SNR	0.962	0.378	0.473	0.811	0.216
NMSE	**0.851**	1.096	1.003	1.094	1.278
ρ	**0.663**	0.579	0.591	0.578	0.422
Without W&G Preprocessing					
	ICA-SCS	Wavelet Shrinkage	Adaptive 2-D Wiener	L2-mean filter	Speckle Image
SNR	**1.002**	0.445	0.535	0.473	0.234
NMSE	1.188	1.198	1.179	1.255	1.397
ρ	0.511	0.573	0.577	0.537	0.448

Table 5.1: Performance measures computed for numerical Exp. No. 1. The values for the original speckle image are listed for reference. The best performance measures have been typed in boldface.

With W&G Preprocessing					
	ICA-SCS	Wavelet Shrinkage	Adaptive 2-D Wiener	L2-mean filter	Speckle Image
SNR	**0.393**	0.127	0.157	0.178	0.059
NMSE	**1.358**	1.550	1.486	1.433	1.749
ρ	**0.479**	0.441	0.433	0.384	0.307
Without W&G Preprocessing					
	ICA-SCS	Wavelet Shrinkage	Adaptive 2-D Wiener	L2-mean filter	Speckle Image
SNR	0.228	0.088	0.104	0.111	0.027
NMSE	1.630	1.641	1.623	1.714	1.790
ρ	0.362	0.421	0.415	0.333	0.335

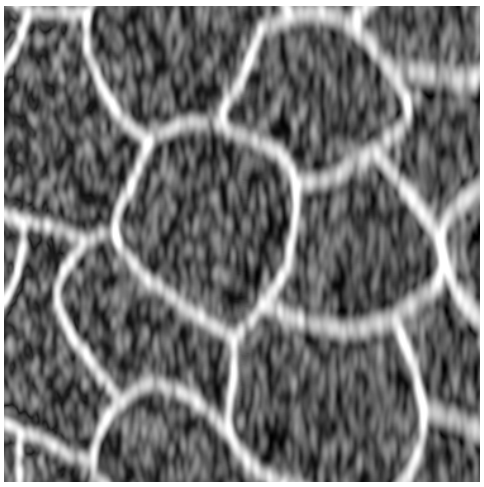
Table 5.2: Performance measures computed for numerical Exp. No. 2. The values for the original speckle image are listed for reference. The best performance measures have been typed in boldface.

With W&G Preprocessing					
	ICA-SCS	Wavelet Shrinkage	Adaptive 2-D Wiener	L2-mean filter	Speckle Image
SNR	1.137	0.535	0.670	1.037	0.312
NMSE	**0.792**	0.979	0.895	1.103	1.157
ρ	**0.675**	0.625	0.634	0.591	0.450
Without W&G Preprocessing					
	ICA-SCS	Wavelet Shrinkage	Adaptive 2-D Wiener	L2-mean filter	Speckle Image
SNR	**1.203**	0.649	0.784	0.617	0.374
NMSE	1.057	1.065	1.053	1.234	1.249
ρ	0.535	0.622	0.626	0.559	0.490

Table 5.3: Performance measures computed for numerical Exp. No. 3. The values for the original speckle image are listed for reference. The best performance measures have been typed in boldface.

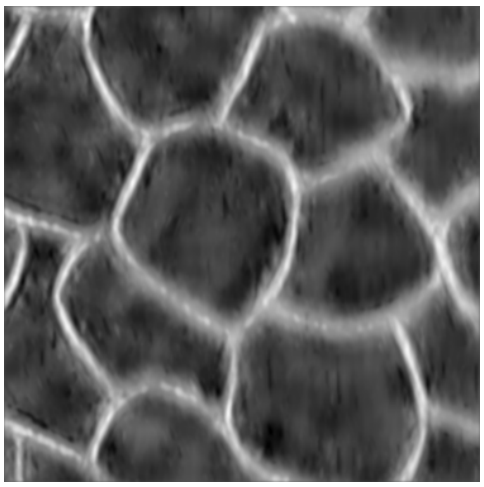
With W&G Preprocessing					
	ICA-SCS	Wavelet Shrinkage	Adaptive 2-D Wiener	L2-mean filter	Speckle Image
SNR	**0.569**	0.254	0.319	0.417	0.111
NMSE	**1.137**	1.300	1.211	1.299	1.589
ρ	**0.536**	0.516	0.513	0.487	0.322
Without W&G Preprocessing					
	ICA-SCS	Wavelet Shrinkage	Adaptive 2-D Wiener	L2-mean filter	Speckle Image
SNR	0.522	0.245	0.301	0.301	0.113
NMSE	1.370	1.393	1.365	1.535	1.606
ρ	0.373	0.508	0.510	0.438	0.376

Table 5.4: Performance measures computed for numerical Exp. No. 4. The values for the original speckle image are listed for reference. The best performance measures have been typed in boldface.

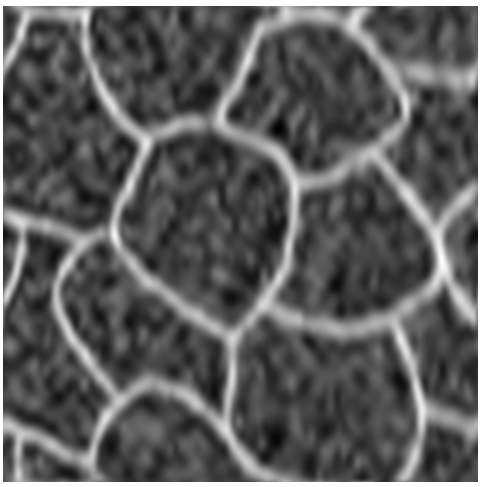


(a) Speckle Image

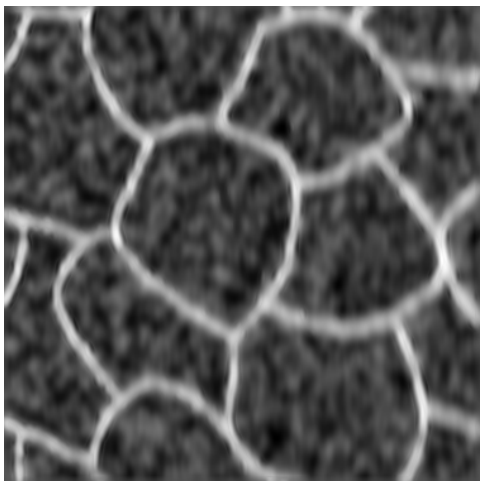
(b) Ground Truth Image



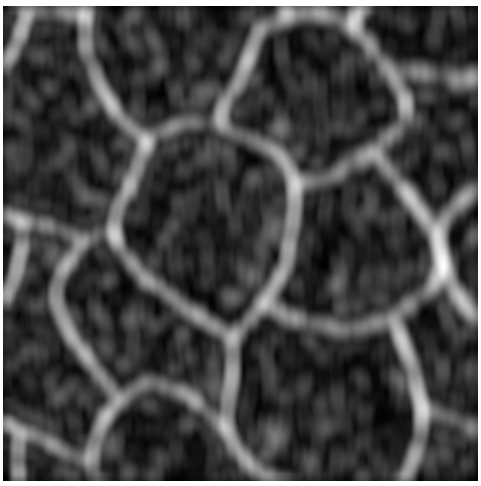
(c) ICA-SCS



(d) Wavelet shrinkage

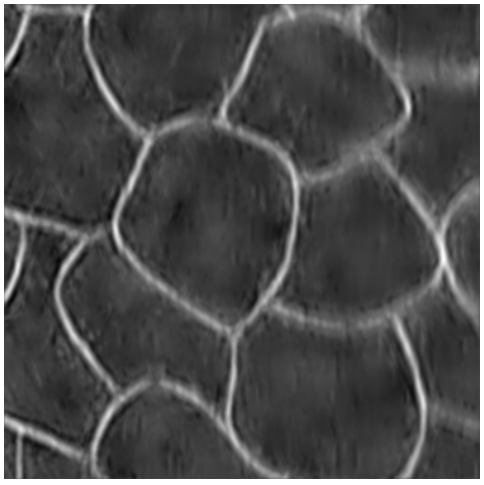


(e) Adaptive 2-D Wiener filter

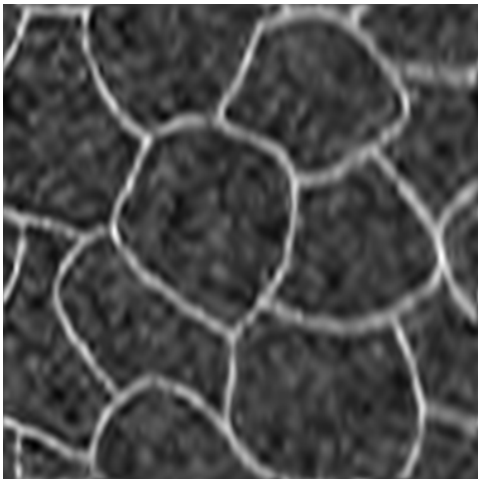


(f) L2-mean filter

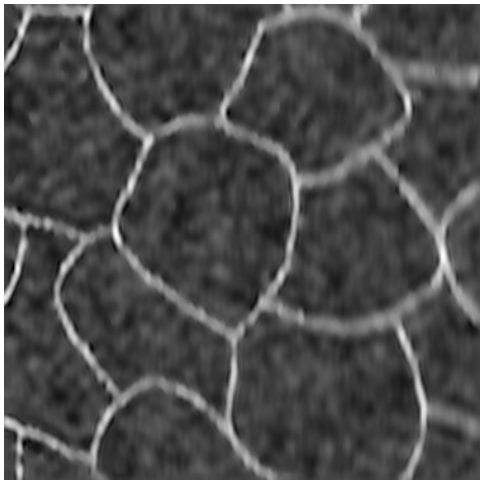
Figure 5.11: Exp. No. 1: (a) Original speckle image, (b) Ground truth image, (c) to (f): The despeckled images using different algorithms without W&G preprocessing step.



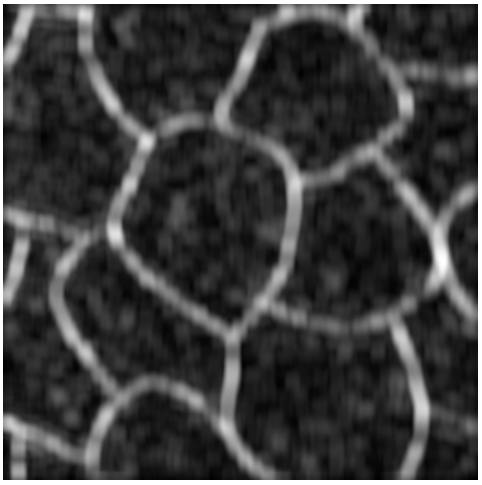
(a) ICA-SCS



(b) Wavelet shrinkage

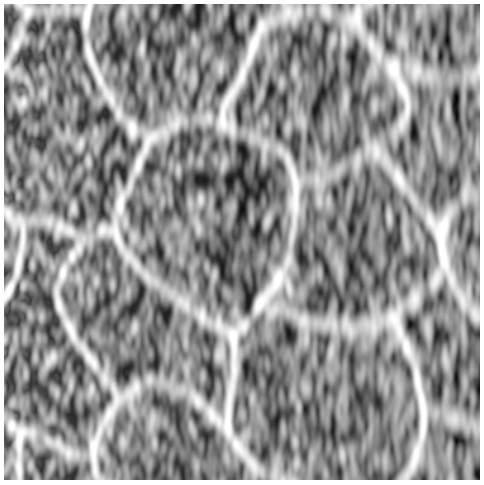


(c) Adaptive 2-D Wiener filter



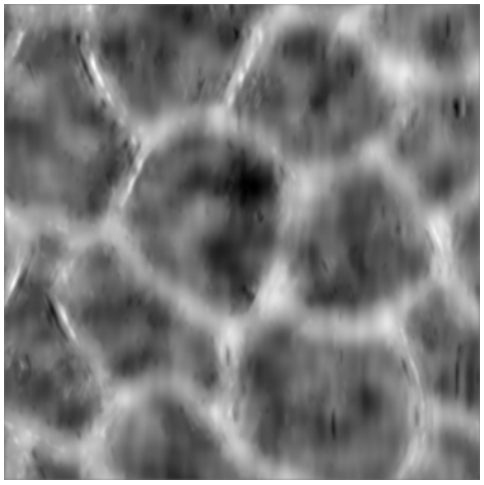
(d) L2-mean filter

Figure 5.12: Exp. No. 1 (cont'd): Despeckled images using different algorithms with W&G preprocessing step. The original speckle image is (a) in Fig. 5.11.

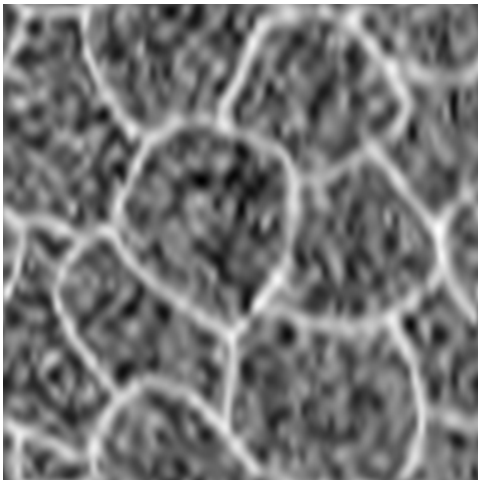


(a) Speckle Image

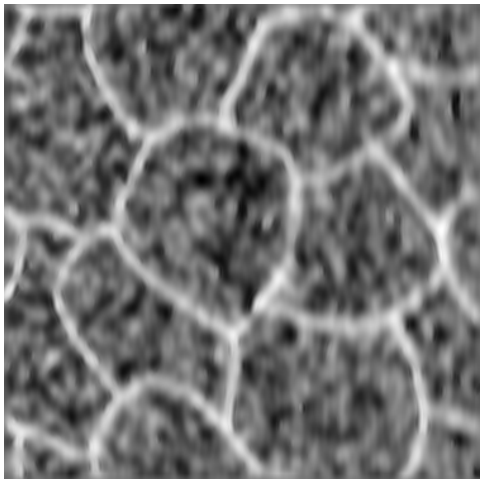
(b) Ground Truth Image



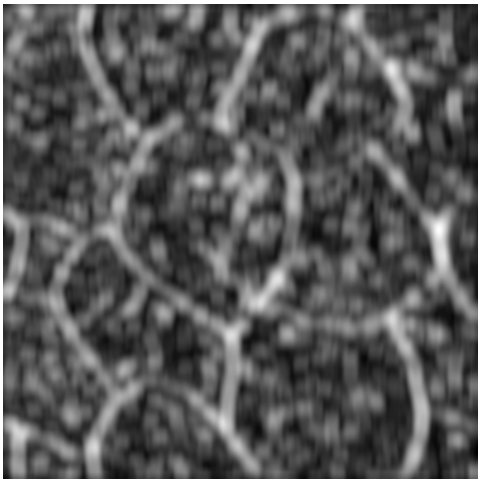
(c) ICA-SCS



(d) Wavelet shrinkage

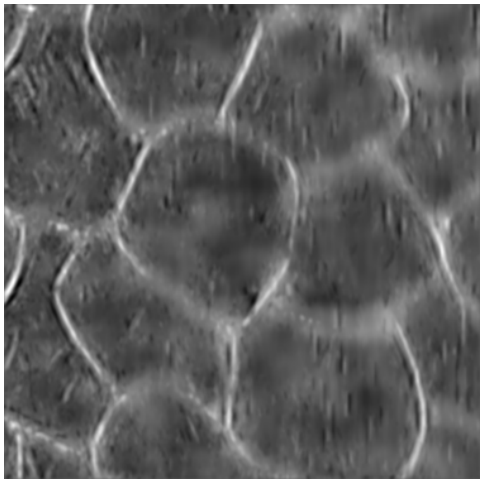


(e) Adaptive 2-D Wiener filter

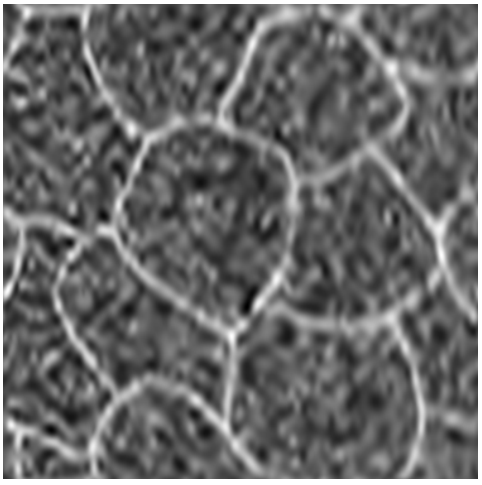


(f) L2-mean filter

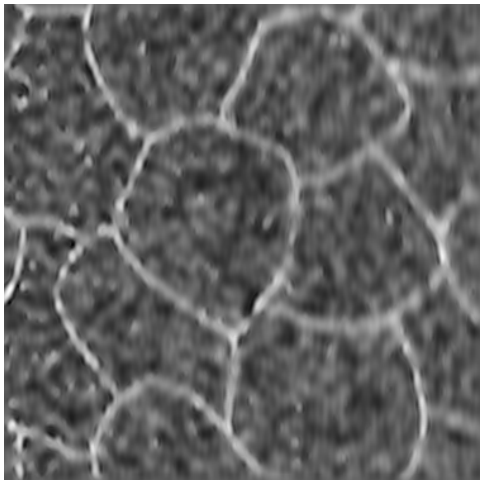
Figure 5.13: Exp. No. 2: (a) Original speckle image, (b) Ground truth image, (c) to (f): The despeckled images using different algorithms without W&G preprocessing step.



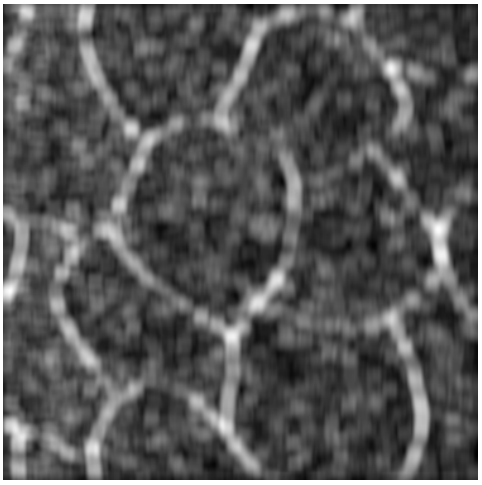
(a) ICA-SCS



(b) Wavelet shrinkage

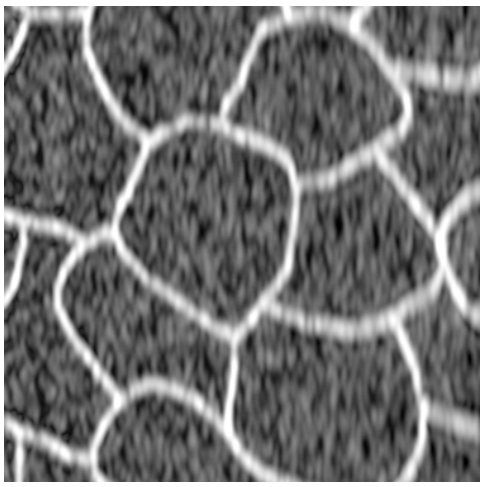


(c) Adaptive 2-D Wiener filter



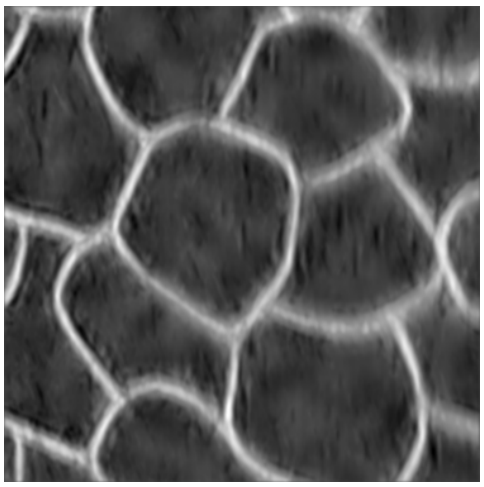
(d) L2-mean filter

Figure 5.14: Exp. No. 2 (cont'd): Despeckled images using different algorithms with W&G preprocessing step. The original speckle image is (a) in Fig. 5.13.

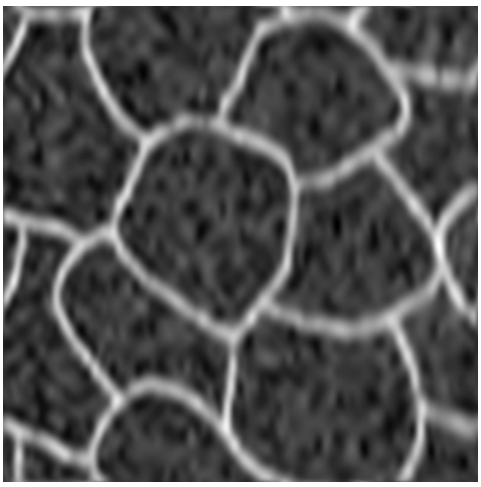


(a) Speckle Image

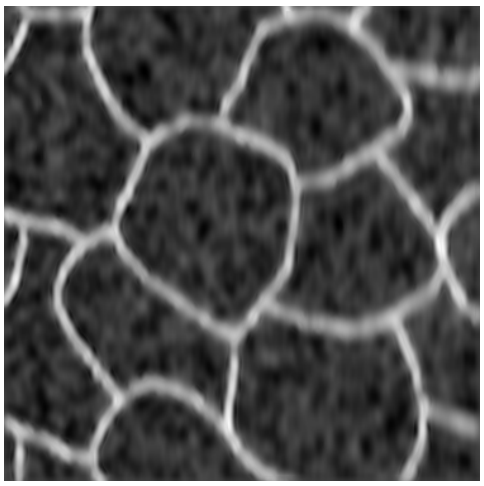
(b) Ground Truth Image



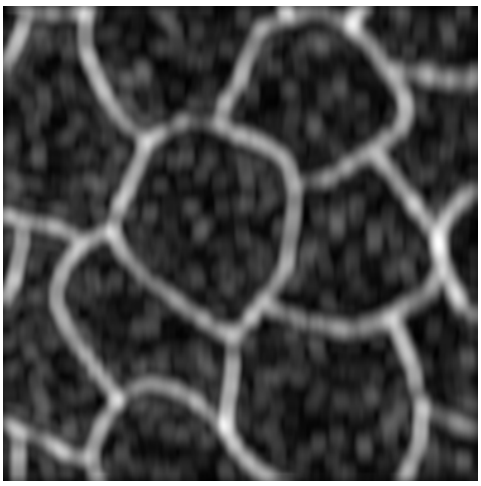
(c) ICA-SCS



(d) Wavelet shrinkage

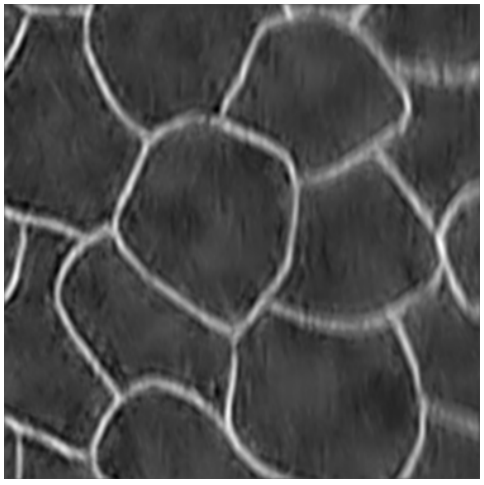


(e) Adaptive 2-D Wiener filter

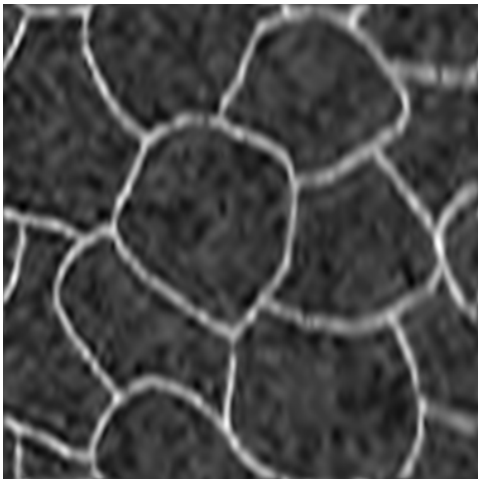


(f) L2-mean filter

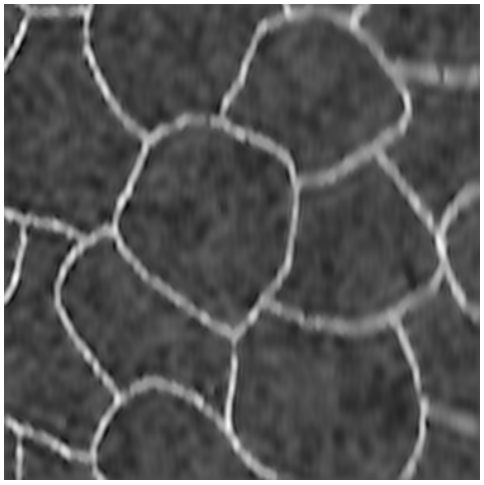
Figure 5.15: Exp. No. 3: (a) Original speckle image, (b) Ground truth image, (c) to (f): The despeckled images using different algorithms without W&G preprocessing step.



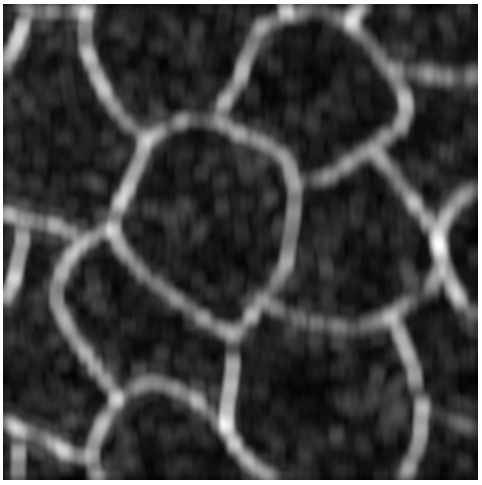
(a) ICA-SCS



(b) Wavelet shrinkage

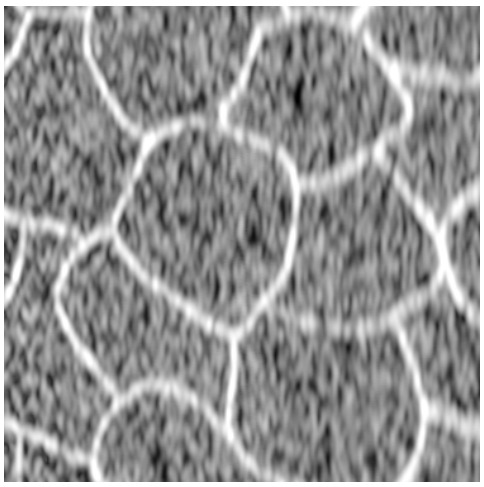


(c) Adaptive 2-D Wiener filter



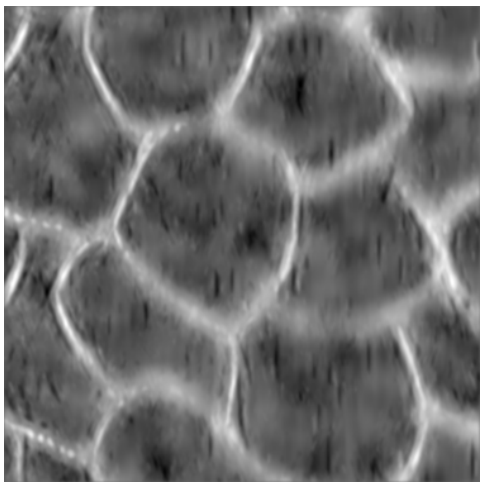
(d) L2-mean filter

Figure 5.16: Exp. No. 3 (cont'd): Despeckled images using different algorithms with W&G preprocessing step. The original speckle image is (a) in Fig. 5.15.

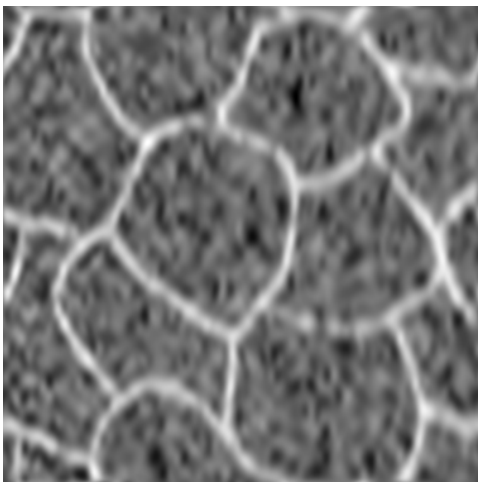


(a) Speckle Image

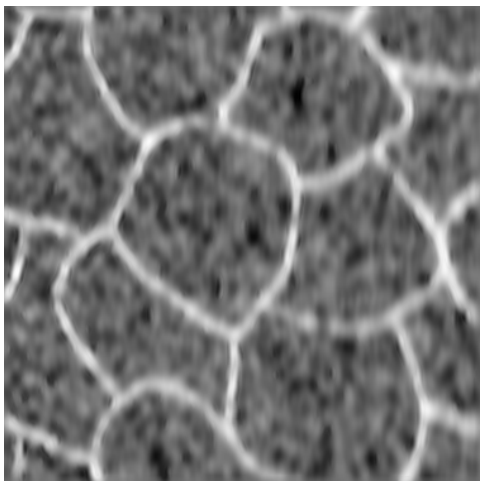
(b) Ground Truth Image



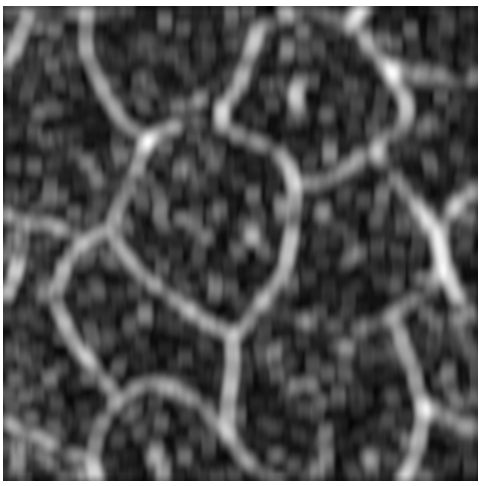
(c) ICA-SCS



(d) Wavelet shrinkage

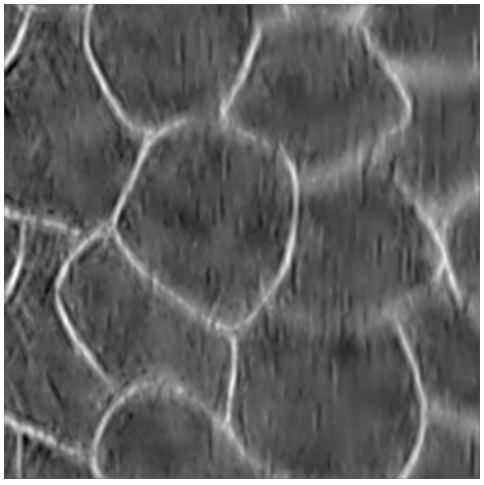


(e) Adaptive 2-D Wiener filter

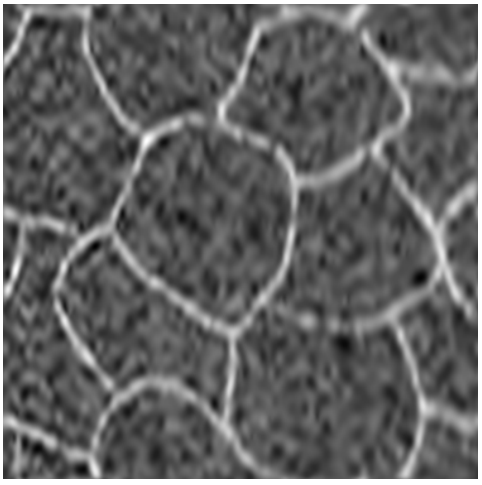


(f) L2-mean filter

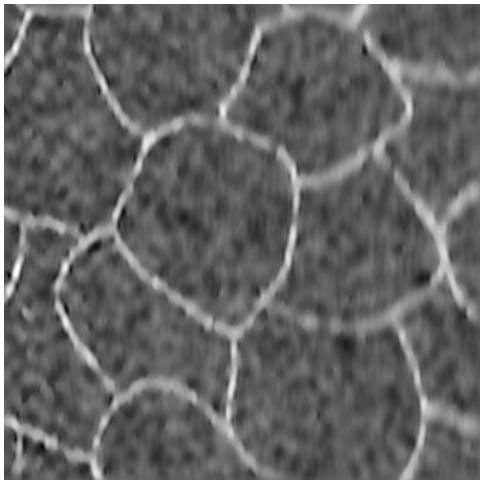
Figure 5.17: Exp. No. 4: (a) Original speckle image, (b) Ground truth image, (c) to (f): The despeckled images using different algorithms without W&G preprocessing step.



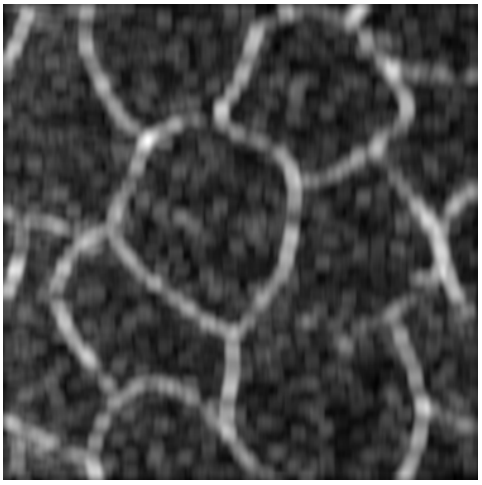
(a) ICA-SCS



(b) Wavelet shrinkage

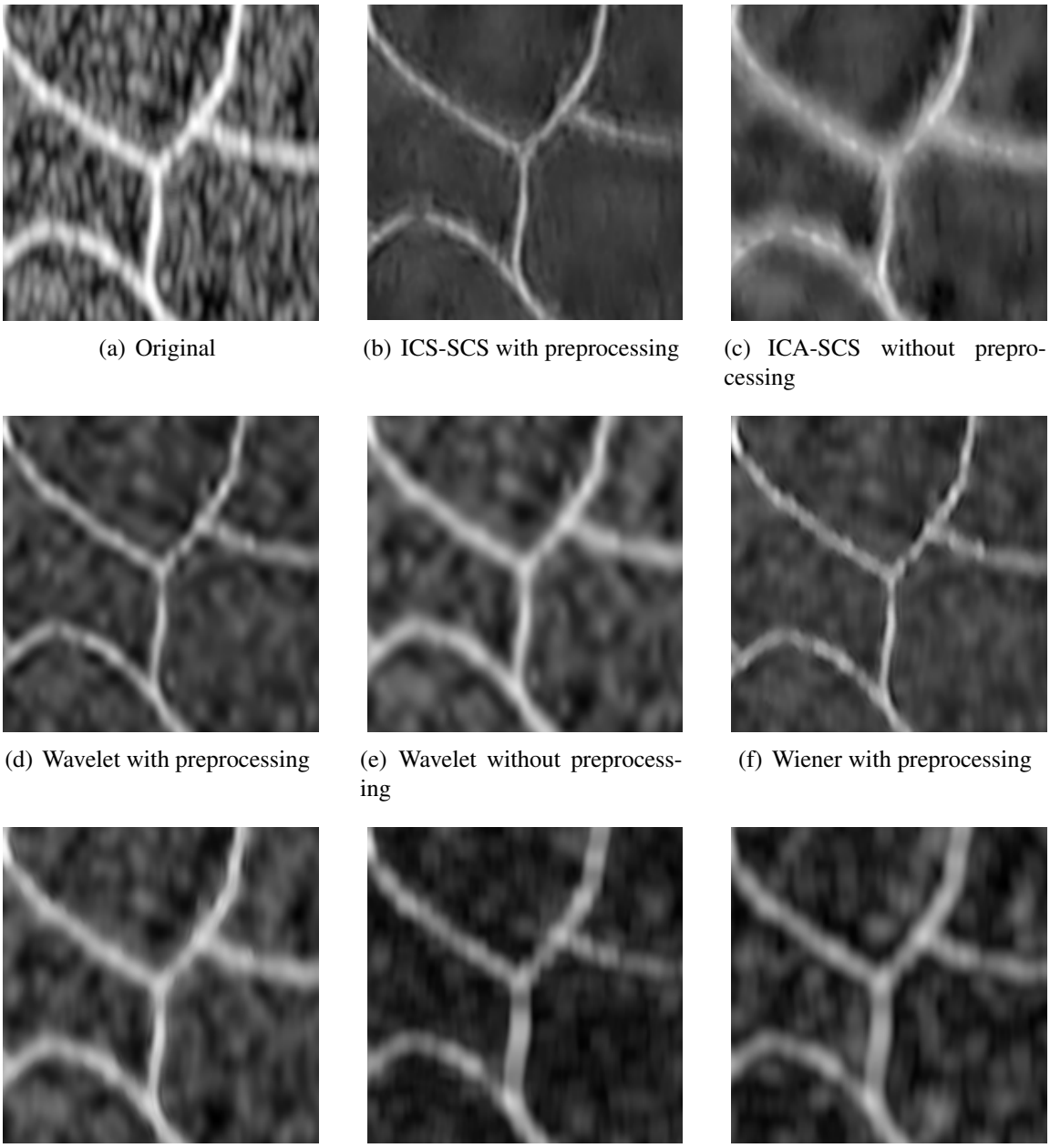


(c) Adaptive 2-D Wiener filter



(d) L2-mean filter

Figure 5.18: Exp. No. 4 (cont'd): Despeckled images using different algorithms with W&G preprocessing step. The original speckle image is (a) in Fig. 5.17.



(a) Original (b) ICS-SCS with preprocessing (c) ICA-SCS without preprocessing

(d) Wavelet with preprocessing (e) Wavelet without preprocessing (f) Wiener with preprocessing

(g) Wiener without preprocessing (h) L2-mean with preprocessing (i) L2-mean without preprocessing

Figure 5.19: A detailed local view for Exp. No. 1. (a) Original (b)-(k) Despeckled images using different methods.

5.8 Discussions

Both the performance measures and the visual appearance of the despeckled images have shown that the proposed ICA-based algorithm is the obvious winner among all the algorithms compared. The three performance measures for the proposed algorithm are always the best ones in each of the four experiments, with only two exceptions (one in Exp. No. 1 and another in Exp. No. 3) that the SNR values of ICA-SCS without preprocessing are slightly larger than that of the proposed algorithm. But the corresponding NMSE values show that the proposed algorithm performs much better than ICA-SCS without preprocessing. Since NMSE is a more reliable measure of speckle reduction than SNR, as we pointed out previously, even in these two exceptional cases the proposed algorithm is still the best performer.

It is observed that the proposed algorithm can effectively reduce the speckle noise while keeping the sharp details of the signal components. ICA-SCS without preprocessing also did a good job in reducing the speckle noise, but the sharp edges were seriously oversmoothed. Generally, when the preprocessing step was applied, all the algorithms did a better job than without the preprocessing step.

The overall performances of wavelet shrinkage and adaptive Wiener filtering are in the middle of the performance rankings for all the algorithms. L2-mean filtering seems to be the worst performer of all. When the speckle noise level tends to be higher as in Exp. No. 2 and No. 4, the performances of the three algorithms for comparison appear to be very weak in both speckle reduction and edge preservation. In the meantime, the proposed algorithm can still achieve fair amount of speckle reduction while keeping some of the sharp edge details. Although one can tune the algorithms for comparison to achieve approximately the same amount of speckle reduction as the proposed method, their performance in edge preservation becomes totally unacceptable. This indicates that since the ICA-based method is high order statistics oriented and more adaptive to the shape of the signal components (in our case, the thin line segments of different angles at the scale of the sampling window),

the signal components can be better differentiated from the speckle components when compared with other algorithms, especially when the noise level becomes higher. For the two different speckle types tested, the proposed algorithm performs better when the speckles are less fully developed. This is due to the fact that the speckles with coherent components are closer to Gaussian noise than fully developed speckles, thus the former suffer less from the outlier problem.

Fully comparing the performance of the proposed algorithm with a wider range of advanced algorithms is beyond the scope of this study. Actually, it seems that there is no single denoising algorithm that claimes to be the best solution for any type of application. Our study has to some extent demonstrated that the proposed algorithm is a very promising despeckling approach specific to ultrasound B-mode images.

The advantage of the proposed algorithm lies in the fact that it considers both high order statistics and data adaptation, as well as the properties of ultrasound speckle. The limitations of the algorithm are that the learned basis functions can only be used for a specific application and enough speckle-free learning samples must be collected to accurately reflect the statistical properties of the signal components in the images for despeckling.

Chapter 6

Contributions, Conclusions and Recommendations

6.1 Contributions

1. To the best of the author's knowledge, this study is the first attempt of applying ICA to B-mode ultrasound speckle texture analysis and tissue characterization.

2. A complete ICA-based speckle texture analysis algorithm, considering the nature of ultrasound images, has been proposed for tissue classification and texture-based segmentation. By comparing the proposed algorithm with several conventional ones, this study has shown that higher order statistics have advantages over first or second order statistics in some situations. It has also shown that ICA filters are more data-adaptive than the predefined filters.

3. Through simulation of speckle textures with underlying tissue properties changing gradually, this study used the proposed texture analysis algorithm as a means of investigating how the behavior of a high order statistics method like ICA changes when the speckle approaches the Rayleigh model from different non-Rayleigh models. It is well known that the

classification efficiency of FOS and SOS drops as speckle approaches the Rayleigh limit. Similar behavior was also observed for ICA.

4. Through the simulation of speckle textures with underlying tissue properties changing gradually, this study showed that the proposed texture analysis algorithm can be potentially useful to the early detection of soft tissue disease.

5. This study proposed a simulation method of controlling the spatial regularity (or randomness) of scatterers with a single parameter using the Poisson disk process. The spatial distribution of a tissue scatterer field can be simulated in the range from full randomness to strong regularity (isotropically).

6. This study is also the first attempt of applying ICA to the B-mode ultrasound despeckling problem. The proposed algorithm combines two existing algorithms with some modifications. Satisfactory results have shown the proposed algorithm to be a promising novel despeckling technique for ultrasound.

6.2 Conclusions

Due to the complicated underlying physical process, it is very difficult to find an analytical physical model to fully describe the ultrasound speckle textures. Even for the first order statistics, there are no perfect universal models to explain a wide range of speckle phenomena. Instead of relying on the model-based approaches, the ICA-based image processing approach provides us an alternative way of dealing with speckle. The ICA-based approach is essentially a blind signal processing approach which does not require much information of the imaging process. It only assumes that the speckle texture pattern is formed as the linear combinations of some basic building blocks, i.e., basis functions.

This study has shown that ICA could be used as a novel statistical tool to tackle the complexity of ultrasound speckle. The proposed ICA-based texture analysis algorithm has shown to be a better performer in many cases than some conventional algorithms. The advantage of the algorithm lies in the fact that it considers both high order statistics and data adaptation.

When the proposed texture analysis algorithm is applied to the speckle textures with underlying tissue properties changing gradually, this study showed how the behavior of high order statistics changes when the speckle approaches the Rayleigh model from different non-Rayleigh models. It has been demonstrated that for the speckle in the limiting case of the Rayleigh model, both the second and high order statistics approaches lost capability of tissue differentiation. However, when the speckles gradually deviate from the Rayleigh model to some non-Rayleigh models, the high order statistics based approach (i.e., the ICA-based approach) shows strong tissue differentiation capability. Whenever there are some spatial regularities added to the tissue, the performance of the ICA-based approach increases quickly. This study also suggested the potential of applying the proposed algorithm to the early detection of soft tissue disease.

The proposed ICA-based texture analysis algorithm has also shown strong capability of multi-class speckle texture segmentation based only on the texture information without utilizing any first order statistics information up to second moments (mean and variance). This is practically significant since sometimes the mean and variance are trivial for tissue differentiation.

For the despeckling problem, the proposed method has shown it to be a strong competitor to many conventional methods. The speckles are effectively reduced while the edge details are well kept. Here again ICA shows its advantage in dissecting the complicated statistical structures of speckle and reducing the computational load of a high dimensional problem.

6.3 Recommendations

6.3.1 Some Extended Work of Speckle Texture Analysis

As reviewed in Chapter 1, the speckle texture depends on both the tissue properties and the imaging conditions. This study has done some work on correlating the tissue properties to the ICA features. However, one may expect that the imaging conditions can also be correlated to the ICA features. For example, one can change the PSF of the imaging system while fixing the properties of the tissue being imaged. Some ICA features may be more sensitive to the change of tissue microstructure while some others more sensitive to the change of imaging conditions. This provides a possible method for separating the instrument dependent factors from the tissue microstructure dependent factors. Then the tissue can be better characterized.

The speckle analysis can also be extended to the scenario in which a tissue is the spatial mixture (instead of a juxtaposition) of two different components. The two components may have quite different inter-scatterer spacing and quite different periodicity. For example, one component may be the diffusive component and the other may be the quasi-periodic specular component. Theoretically, this is a blind convolution-mixture separation problem which may be very difficult to solve.

This study is based on the envelope detected images. After envelope detection, some extra information contained in the RF signal may be lost. Ideally, if the ICA-based approach can be applied to RF images, one may expect that the performance of texture classification and tissue characterization could be improved. However, the difficulty lies in that for the same size of window, the RF image contains many more samples than the envelope image. Thus the computation load of an ICA algorithm becomes extremely high. If we reduce the window to a rather small size, the ICA-based algorithm may fail to catch the local speckle signature. This issue is worth future investigation.

6.3.2 Despeckling Based on Multiplicative ICA Noise Model

To find a rigorously defined ICA speckle model would be a challenging job. Blanco *et al.* [1] proposed a speckle ICA model (SICA) which assumes that the observed signal is contaminated with multiplicative speckle noise. This can be expressed as

$$z_i = v_i x_i, \qquad i = 1, \cdots, N \quad \text{with} \quad \mathbf{x} = \mathbf{As} \tag{6.1}$$

where $\mathbf{s}$ is the vector of independent sources and $\mathbf{v}$ is the speckle noise vector with unit mean. The components of $\mathbf{v}$ are mutually independent and the noise free data $\mathbf{x}$ and $\mathbf{v}$ are also independent. Given a set of speckle signal $\mathbf{z}$, SICA uses some particular structure that signal $\mathbf{z}$ has in its second and third order statistics. We believe this model can be further applied to the despeckling problem found in coherent imaging modalities such as ultrasound imaging.

6.3.3 Blind Deconvolution Using ICA

Blind deconvolution (BD) is used in ultrasound imaging to restore an image with limited resolution. A lot of research work has been done on this subject. ICA can be used to solve the BD problem from a different perspective. Kaplan *et al.* [43] proposed a BD algorithm using ICA with a banded mixing matrix. This algorithm only deals with 1D signal but we believe it can be possibly extended to solve the image BD problem. A good BD algorithm can improve the resolution of an ultrasound image and may be used as a preprocessing step for despeckling.

6.3.4 Multi-frame Ultrasound Image Processing Using ICA

One of the despeckling approaches used in ultrasound imaging is to scan the same tissue region several times, either from different directions or from the same direction but with different transducer performance characteristics such as frequency bandwidth or center fre-

quency. The obtained multiple images from the same region can then be averaged (compounding), followed by traditional restoration methods, to reduce speckle. Beside noise reduction, using multi-frame (or multichannel) images could also obtain more information about the microstructure of the tissue because the tissue's response to ultrasound may vary at different frames (or bands). A similar idea can be found in remote sensing in which hundreds of narrow band spectral images of the same scene are stacked. A vector can be defined for each pixel across all bands. In ultrasound imaging, Chen *et al.* [9] proposed a new compounding method based on the independent noise constraint. We believe that the idea of ICA can be further applied to this type of multi-frame (or multi-channel) image processing problem.

Appendix A

Proof: Orthogonally Transformed White Signal Is Still White

Assume $\mathbf{z}$ is white, which means that $\mathrm{E}\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$. Now consider an orthogonal transform $\mathbf{U}$ of $\mathbf{z}$:

$$\mathbf{y} = \mathbf{U}\mathbf{z} \tag{A.1}$$

Due to the orthogonality of $\mathbf{U}$, we have

$$\mathrm{E}\{\mathbf{y}\mathbf{y}^T\} = \mathrm{E}\{\mathbf{U}\mathbf{z}\mathbf{z}^T\mathbf{U}^T\} = \mathbf{U}\mathbf{I}\mathbf{U}^T = \mathbf{I} \tag{A.2}$$

This proves that $\mathbf{y}$ is also white.

Appendix B

Proof: The Joint Distribution of a Gaussian Mixture Does Not Change with the Orthogonal Mixing Matrix

Assume both $\mathbf{x}$ and $\mathbf{y}$ are N-dimensional random vectors that are related by the vector mapping

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \tag{B.1}$$

for which the inverse mapping

$$\mathbf{x} = \mathbf{g}^{-1}(\mathbf{y}) \tag{B.2}$$

exists and is unique. Then the density $p_{\mathbf{y}}(\mathbf{y})$ of $\mathbf{y}$ is obtained from the density $p_{\mathbf{x}}(\mathbf{x})$ as follows:

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{1}{|\det J\mathbf{g}(\mathbf{g}^{-1}(\mathbf{y})|} p_{\mathbf{x}}(\mathbf{g}^{-1}(\mathbf{y})) \tag{B.3}$$

where $J\mathbf{g}$ is the Jacobian matrix.

Now assume that the joint distribution of two ICs, s_1 and s_2, is Gaussian. The joint *pdf* is given by

$$p(s_1, s_2) = \frac{1}{2\pi} exp(-\frac{s_1^2 + s_2^2}{2}) = \frac{1}{2\pi} exp(-\frac{\|\mathbf{s}\|^2}{2}) \tag{B.4}$$

Now assume that the mixing matrix $\mathbf{A}$ is orthogonal (this is reasonable because we assume the data have been whitened). Using formula B.3, and also noting that for an orthogonal matrix $\mathbf{A}$, $\mathbf{A}^{-1} = \mathbf{A}^T$ holds. Then we get the joint density of the mixtures x_1 and x_2 given by

$$p(x_1, x_2) = \frac{1}{2\pi} exp(-\frac{||\mathbf{A}^T\mathbf{x}||}{2})|\text{det}\mathbf{A}^T| \tag{B.5}$$

Due to the orthogonality of $\mathbf{A}$, we have $||\mathbf{A}^T\mathbf{x}||^2 = ||x||^2$ and $|\text{det}\mathbf{A}| = 1$. Also note that if $\mathbf{A}$ is orthogonal, so is $\mathbf{A}^T$. Thus we have

$$p(x_1, x_2) = \frac{1}{2\pi} exp(-\frac{||\mathbf{x}||^2}{2}) \tag{B.6}$$

so we can see that the orthogonal mixing matrix does change the *pdf*, since it does not appear in the above *pdf* at all. The original (B.4) and mixed (B.6) distributions are identical. Therefore, there is no way that we could infer the mixing matrix from the mixtures. Please see [30] for further reference.

Appendix C

Fast Fixed-point ICA Algorithm Using Kurtosis

An ordinary gradient algorithm using kurtosis computes the direction in which the absolute value of the kurtosis of $y = \mathbf{w}^T\mathbf{z}$ is growing most strongly, based on the available samples of $\mathbf{z}$. Here $\mathbf{z}$ is a whitened version of observed signal $\mathbf{x}$ ($\mathbf{z} = \mathbf{V}\mathbf{x}$ and $\mathbf{V}$ is a whitening matrix). The practical gradient algorithm can be given as

$$\triangle\mathbf{w} \propto \mathbf{w}(Kurt(\mathbf{w}^T\mathbf{z}))\mathrm{E}\{\mathbf{z}(\mathbf{w}^T\mathbf{z})^3\} \tag{C.1}$$

$$\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\| \tag{C.2}$$

The drawback of this ordinary gradient algorithm is that the convergence is slow and depends on a good choice of the learning rate sequence. A bad choice of the learning rate can destroy the convergence. In order to overcome the drawback of this ordinary gradient algorithm, Hyvärinen, *et al.* [28] designed a new algorithm called FastICA. It is an efficient fixed-point iterative algorithm which is given as

$$\mathbf{w} \leftarrow \mathrm{E}\{\mathbf{z}(\mathbf{w}^T\mathbf{z})^3\} - 3\mathbf{w}] \tag{C.3}$$

The final vector $\mathbf{w}$ gives one of the independent components as the linear combination $\mathbf{w}^T\mathbf{z}$.

The convergence of this algorithm is fast (in *cubic*). Contrary to an ordinary gradient-based algorithm, there is no learning rate or other adjustable parameters, which makes it easy to use and more reliable. The realization of this algorithm in Matlab can be downloaded from the algorithm designer's web site:

`www.cis.hut.fi/projects/ica/fastica`

Appendix D

Envelope Detection of Amplitude Modulated Signal Using Hilbert Transform

Mathematically, the Hilbert transform [25] of a function $f(t)$ can be defined as the convolution of this function with function $h(t) = \frac{1}{\pi t}$

$$H(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{f(\tau)}{t-\tau} \mathrm{d}\tau \tag{D.1}$$

where P is called the Cauchy principal value. In signal processing, an amplitude modulated real signal $u(t)$ can be modeled as the production of a bandlimited "message" waveform $u_m(t)$ and a sinusoidal "carrier" as below

$$u(t) = u_m(t) \cdot \cos(\omega t + \phi) \tag{D.2}$$

When $u_m(t)$ has no frequency content above the carrier frequency $\frac{\omega}{2\pi}$, then the Hilbert transform of $u(t)$ can be expressed as

$$H_u(t) = u_m(t) \cdot \sin(\omega t + \phi) \tag{D.3}$$

So it is clear that here the Hilbert transform shifts 90° of the phase of carrier wave. If the signal is expressed as the so-called *analytical signal* form then we obtain

$$\begin{aligned} u_a(t) &= u(t) + i \cdot H_u(t) \\ &= u_m(t) \cdot e^{i(\omega t + \phi)} \end{aligned} \tag{D.4}$$

The "message" signal (envelope signal) $u_m(t)$ can then be easily obtained by taking the amplitude of the complex analytical signal $u_a(t)$

$$\begin{aligned} u_m(t) &= |u(t) + i \cdot H_u(t)| \\ &= \sqrt{|u(t)|^2 + |H_u(t)|^2} \end{aligned} \tag{D.5}$$

This detected envelope signal is sometimes called in-phase/quadrature (I/Q) signal. The Matlab code used to obtain the envelope signal is given as

```
env=abs(hilbert(x))
```

where x is a RF signal and env is the envelope detected signal.

Appendix E

Complete Experimental Results of Chapter 4

E.1 Results of Exp. No. 1

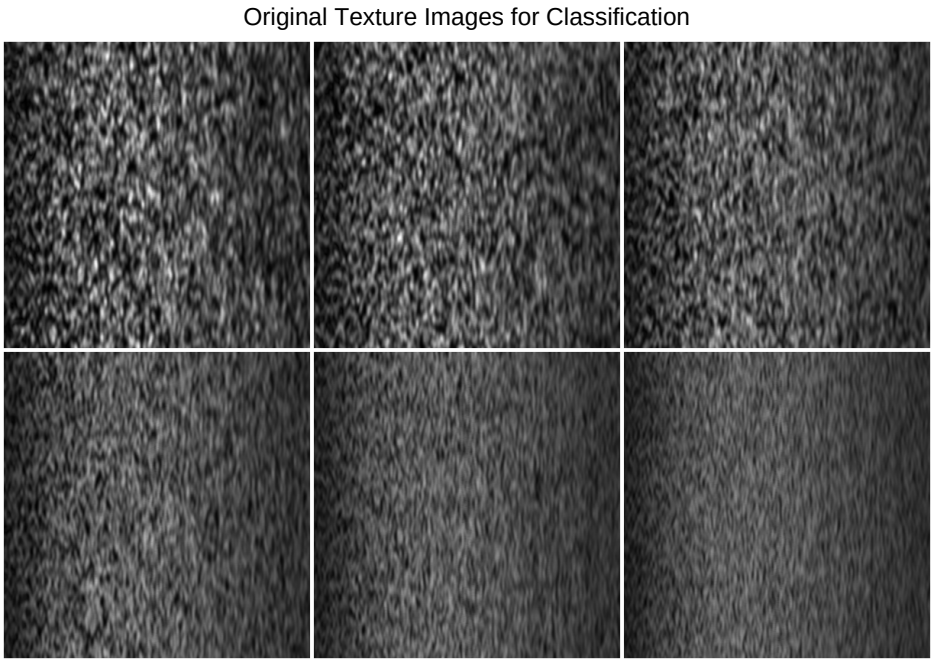


Figure E.1: 6 original B-mode texture images for classification. Top row from left to right: T1-0, T1-1, T1-2; Bottom row from left to right: T1-3, T1-4, T1-5.

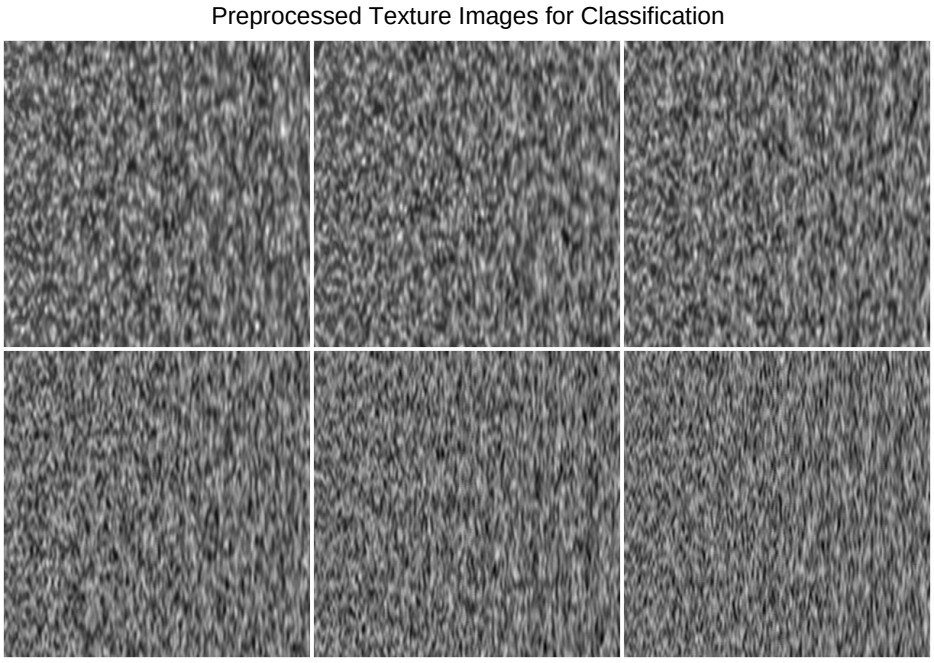
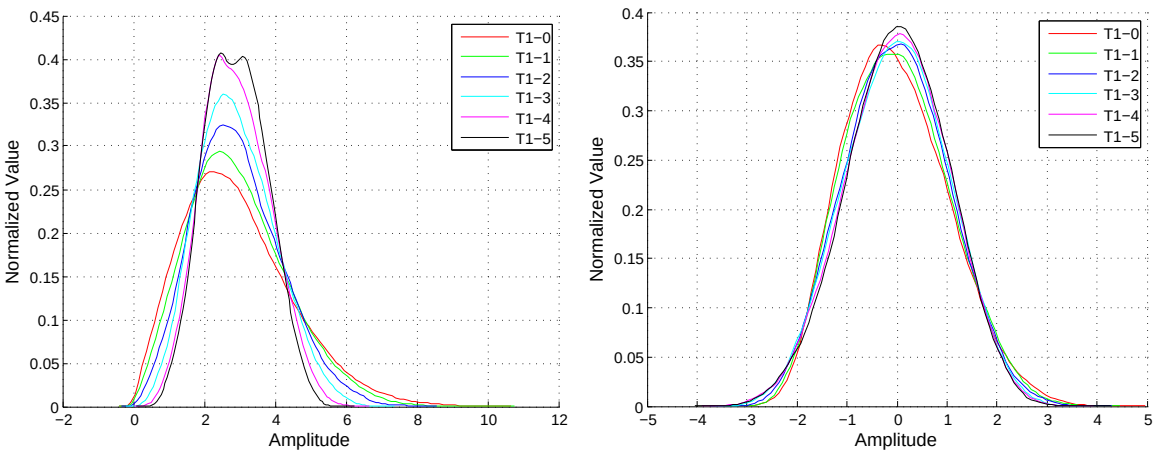


Figure E.2: 6 preprocessed B-mode texture images for classification. Top row from left to right: T1-0, T1-1, T1-2; Bottom row from left to right: T1-3, T1-4, T1-5.



(a) Before preprocessing (b) After preprocessing

Figure E.3: Histogram of the 6 texture images before and after preprocessing.

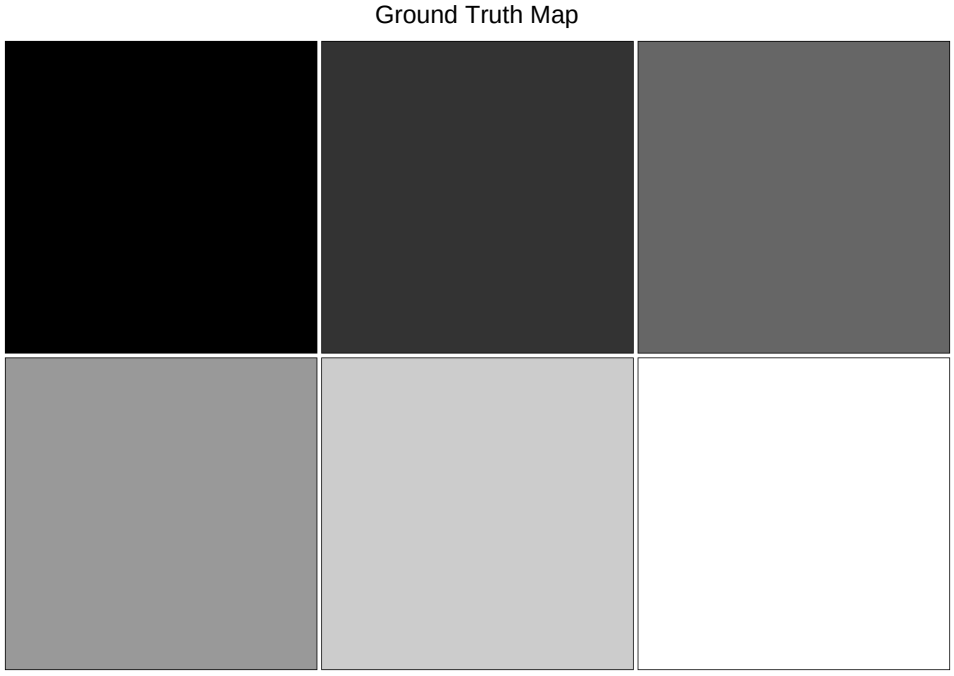


Figure E.4: Corresponding ground truth class map of the 6 texture images.

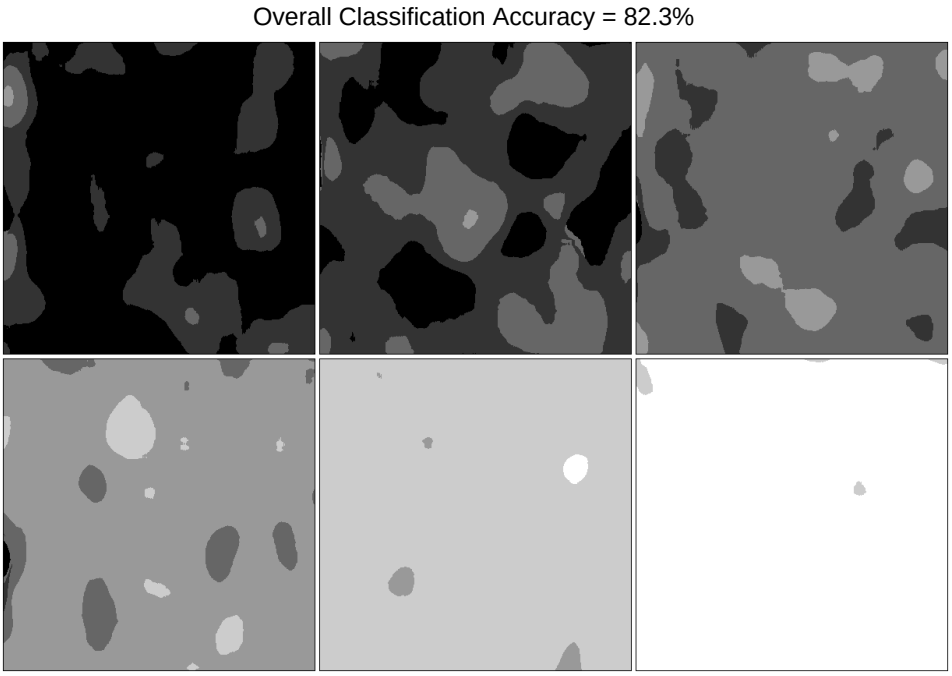


Figure E.5: Classification map using ICA (Exp. No. 1)

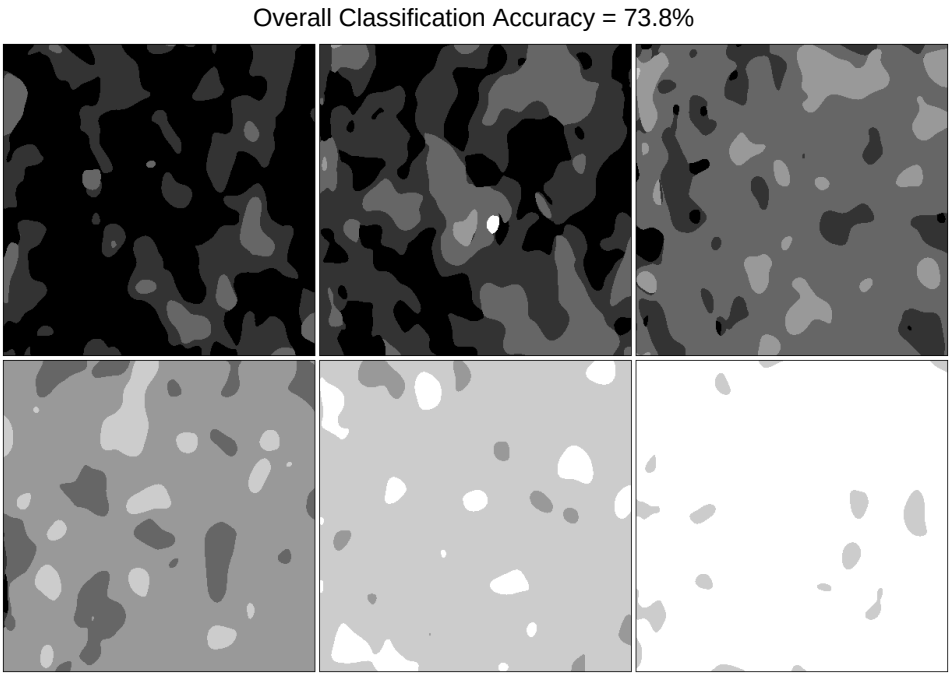


Figure E.6: Classification map using PCA (Exp. No. 1)

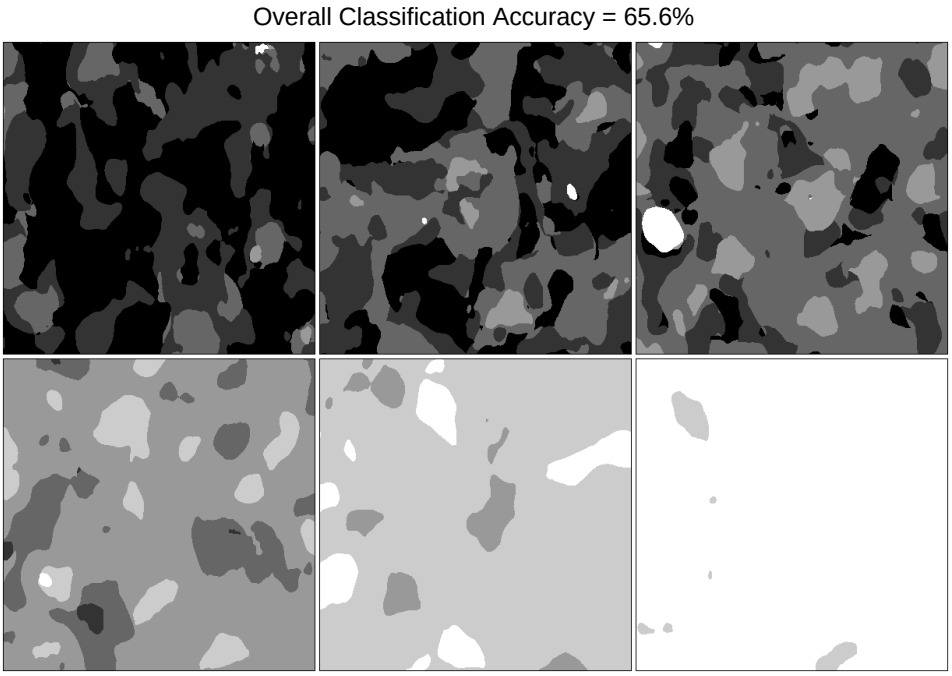


Figure E.7: Classification map using Gabor (Exp. No. 1)

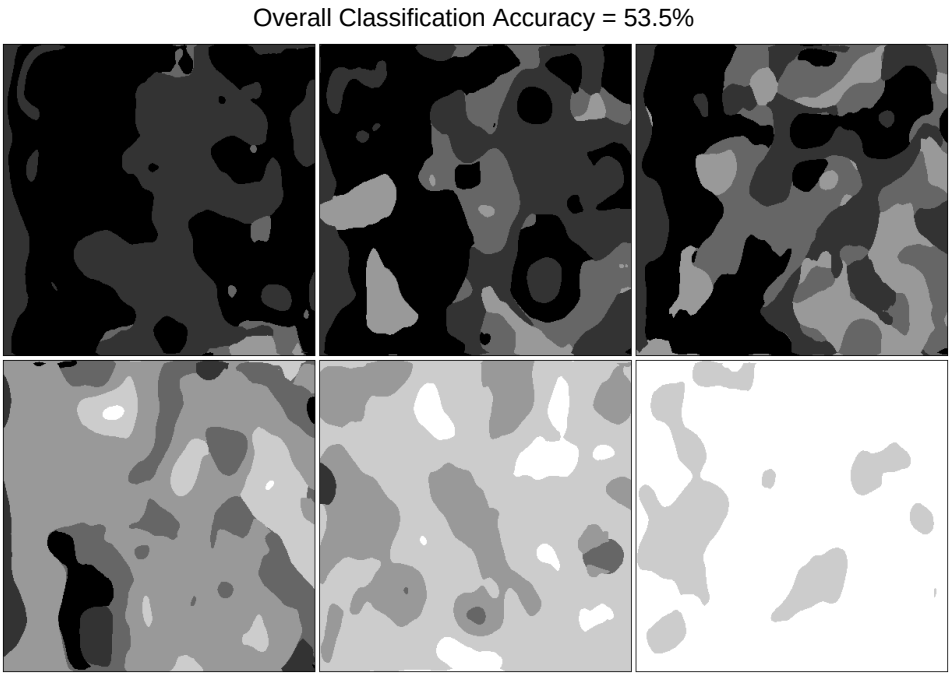


Figure E.8: Classification map using FWHM-ACVF (Exp. No. 1)

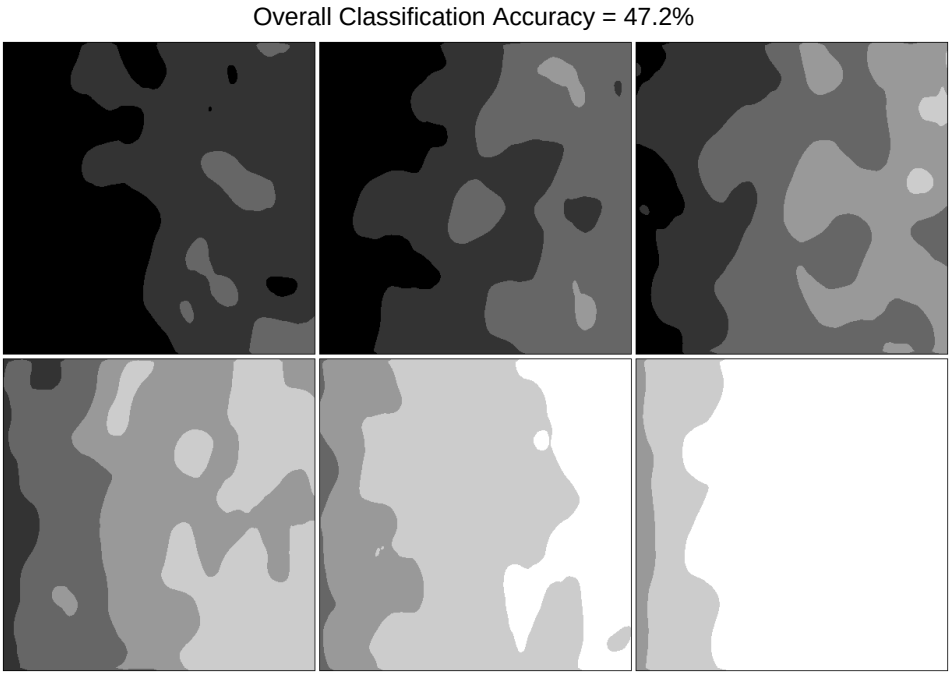


Figure E.9: Classification map using SNR-ALPHA (Exp. No. 1)

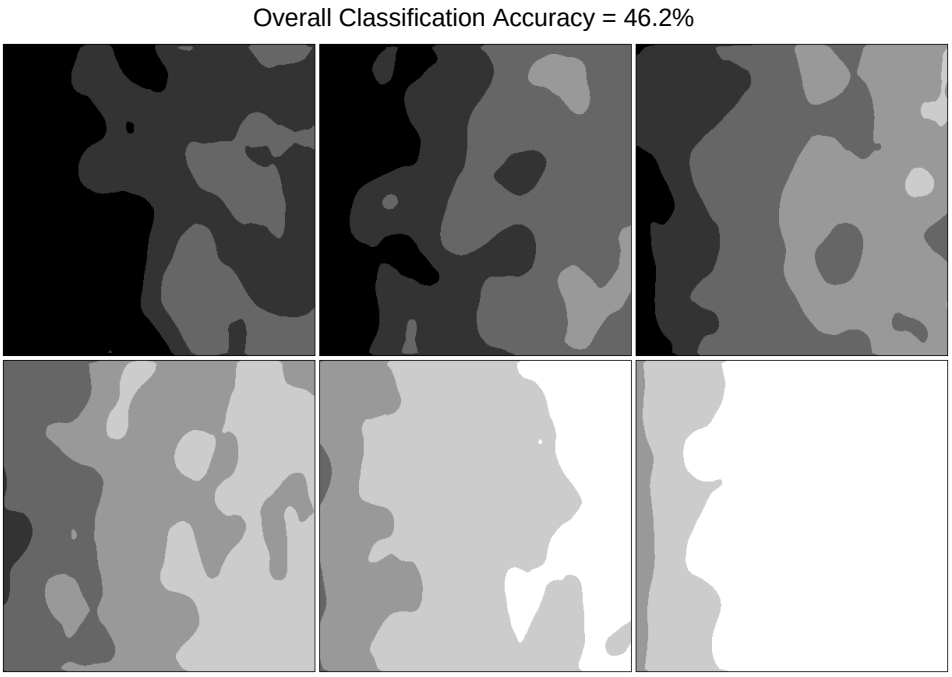
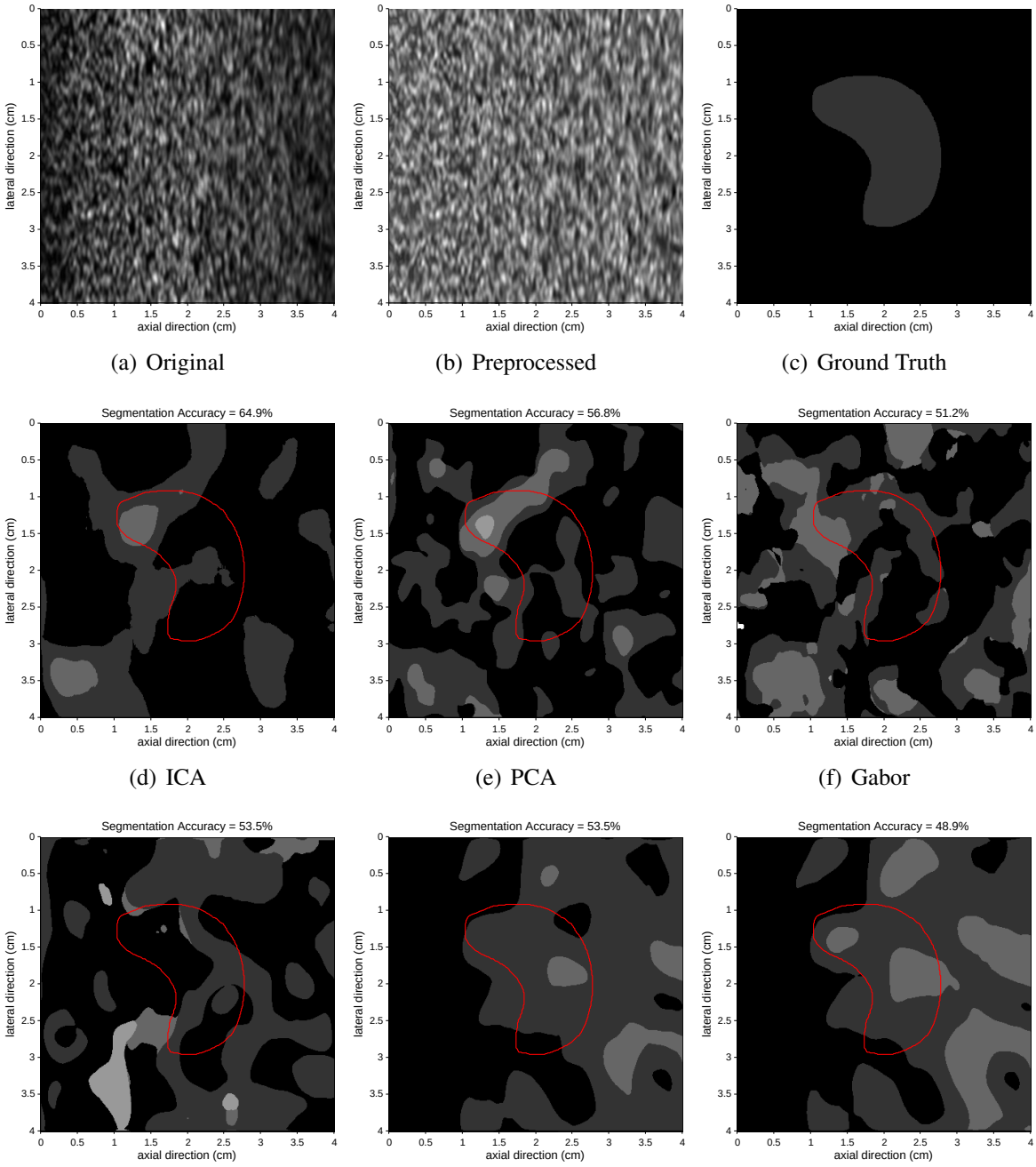


Figure E.10: Classification map using SNR-GAMMA (Exp. No. 1)

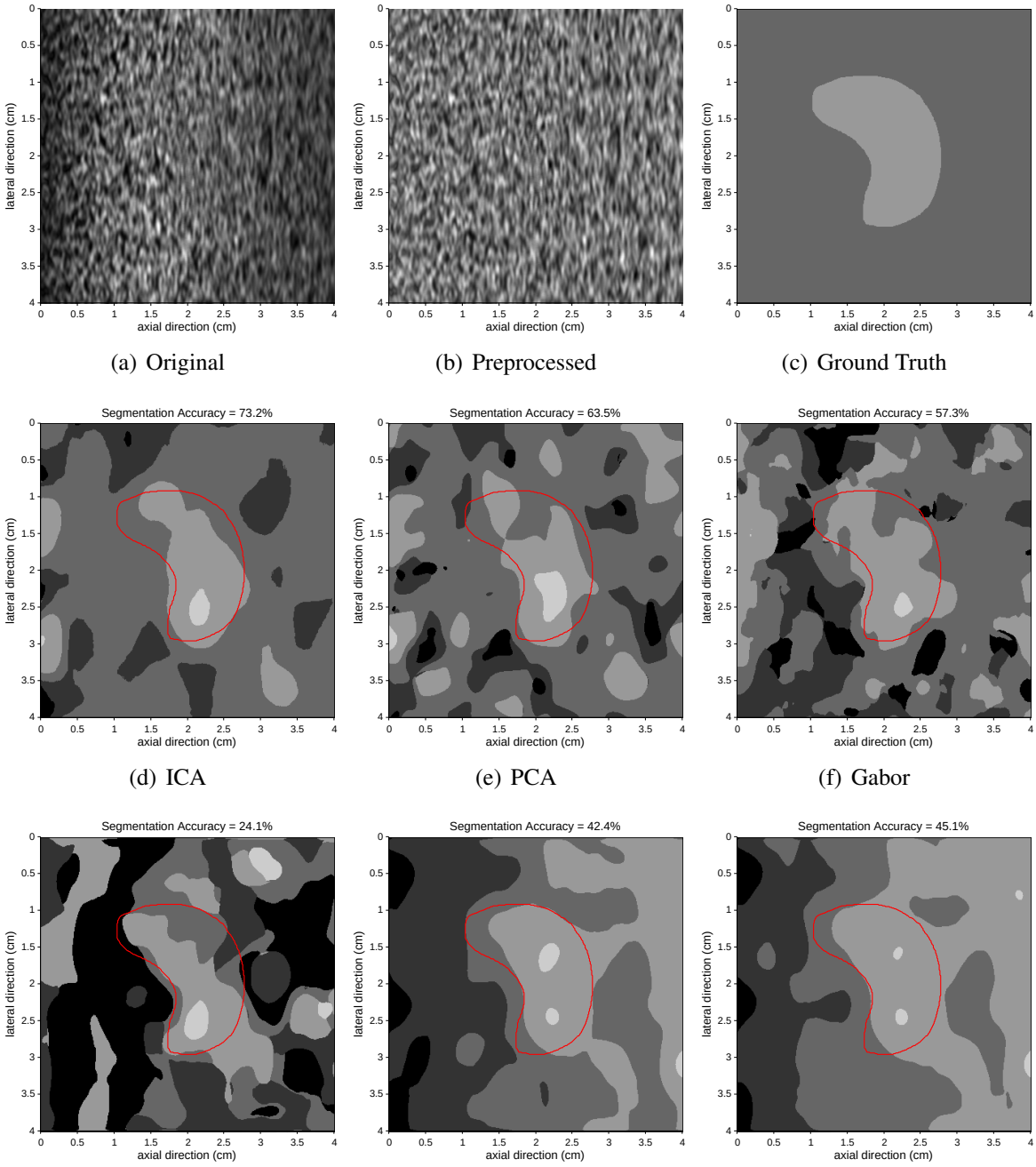


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM-ACVF (h) SNR-ALPHA (i) SNR-GAMMA

Figure E.11: Two-component Texture Segmentation Results (Background:T1-0, Lesion:T1-1)

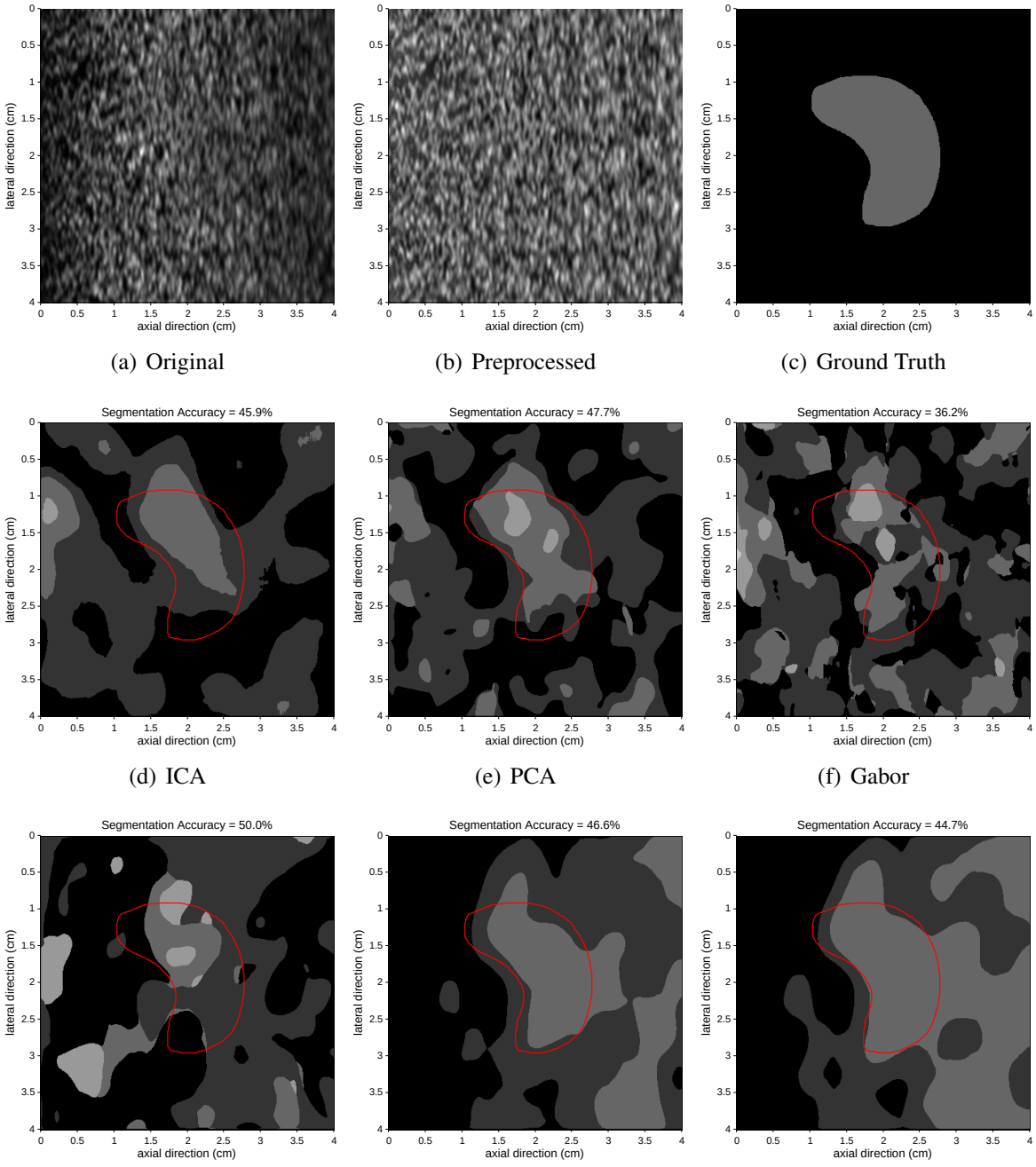


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM-ACVF (h) SNR-ALPHA (i) SNR-GAMMA

Figure E.12: Two-component Texture Segmentation Results (Background:T1-2, Lesion:T1-3)

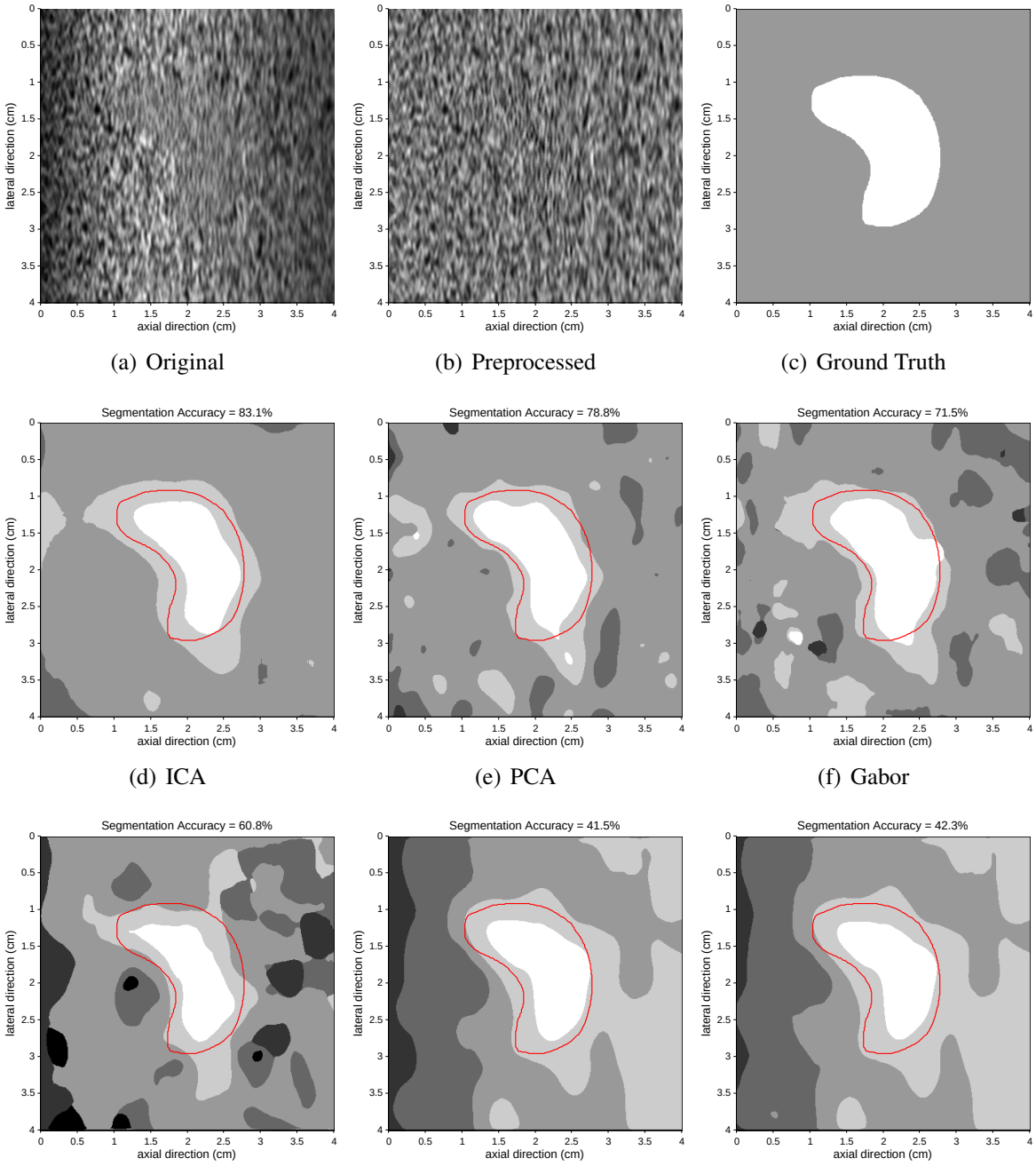


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM-ACVF (h) SNR-ALPHA (i) SNR-GAMMA

Figure E.13: Two-component Texture Segmentation Results (Background:T1-0, Lesion:T1-2)

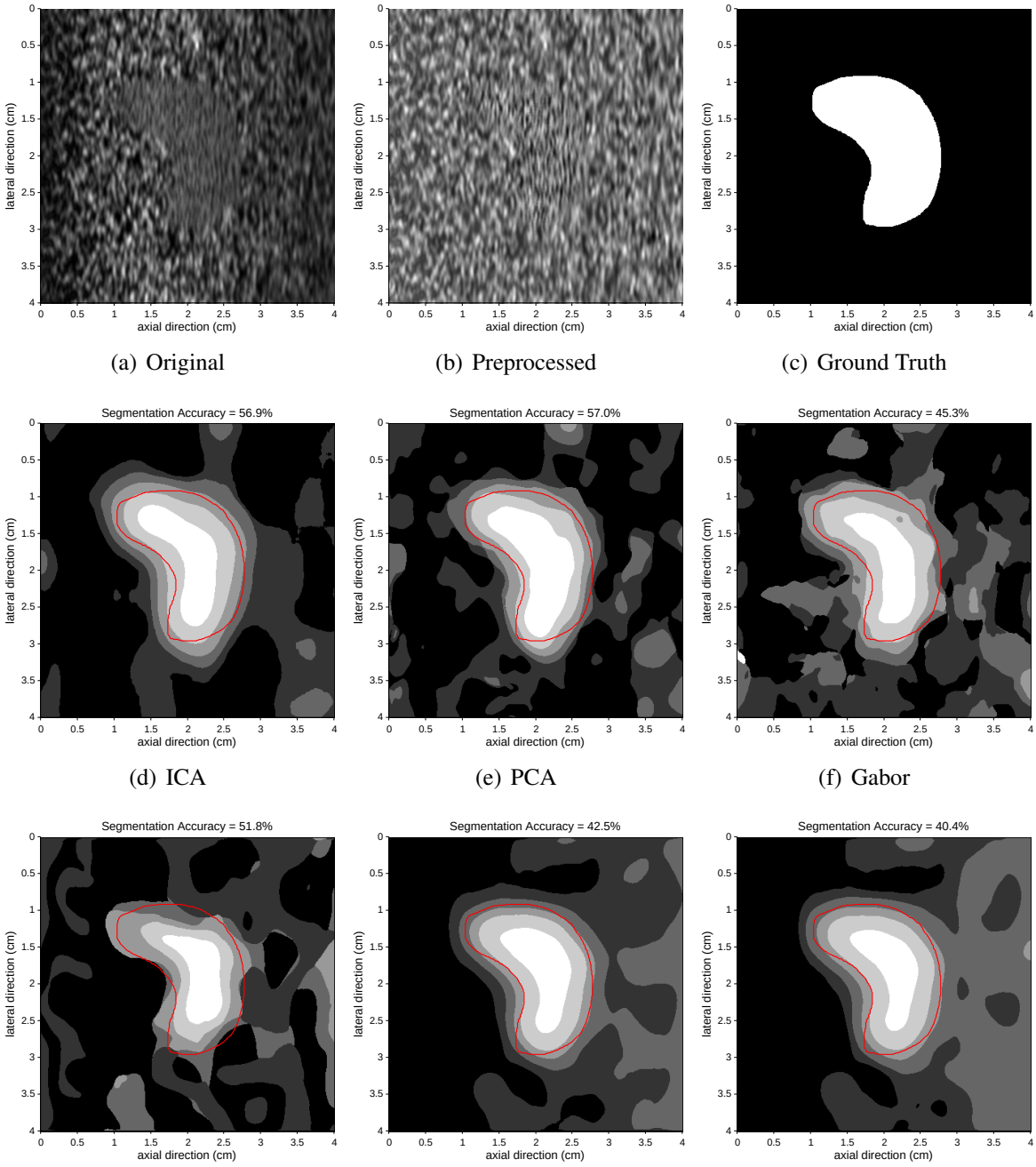


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM-ACVF (h) SNR-ALPHA (i) SNR-GAMMA

Figure E.14: Two-component Texture Segmentation Results (Background:T1-3, Lesion:T1-5)



(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM-ACVF (h) SNR-ALPHA (i) SNR-GAMMA

Figure E.15: Two-component Texture Segmentation Results (Background:T1-0, Lesion:T1-5)

E.2 Results of Exp. No. 2

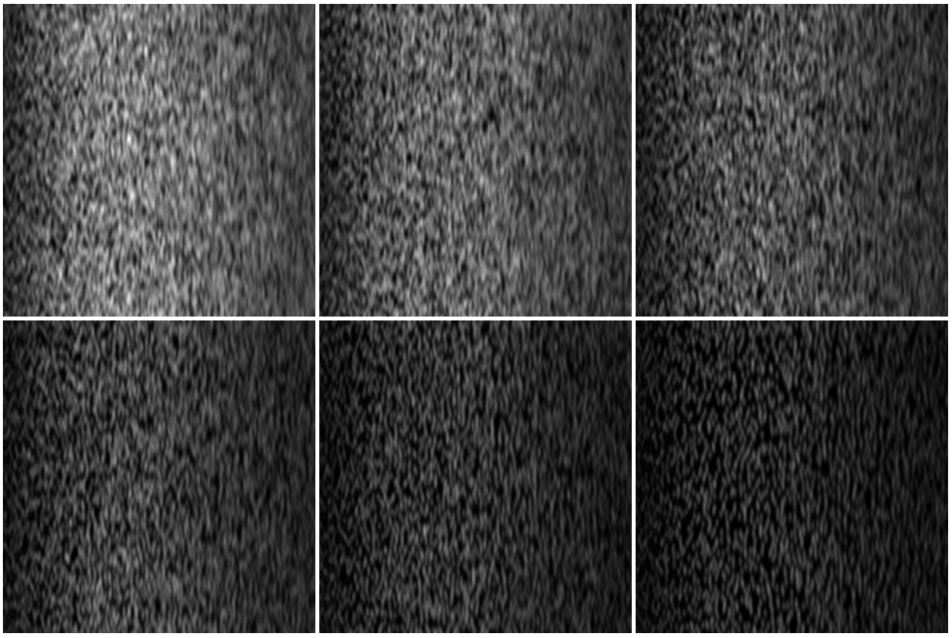
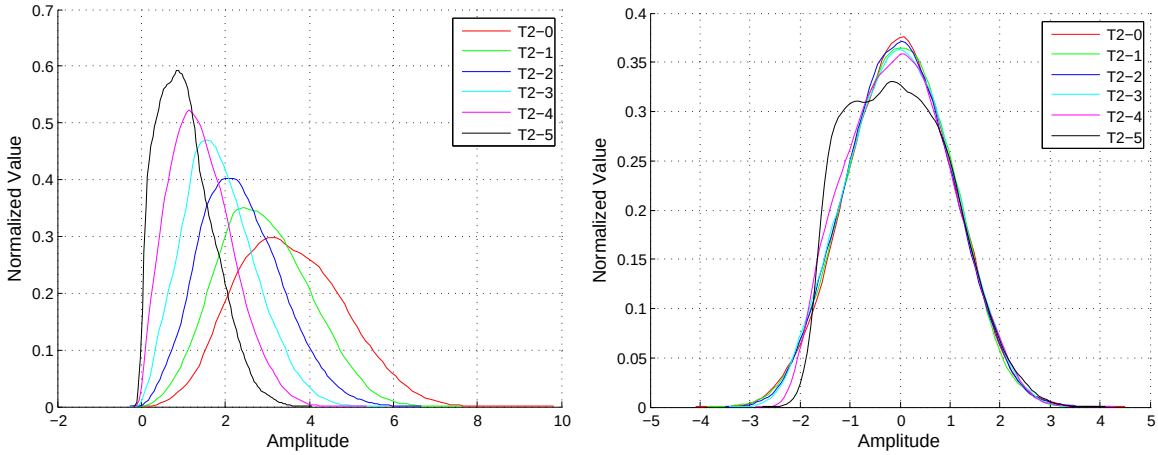


Figure E.16: 6 original B-mode texture images for classification. Top row from left to right: T2-5, T2-4, T2-3; Bottom row from left to right: T2-2, T2-1, T2-0.

Figure E.17: 6 preprocessed B-mode texture images for classification. Top row from left to right: T2-5, T2-4, T2-3; Bottom row from left to right: T2-2, T2-1, T2-0.



(a) Before preprocessing (b) After preprocessing

Figure E.18: Histogram of the 6 texture images before and after preprocessing

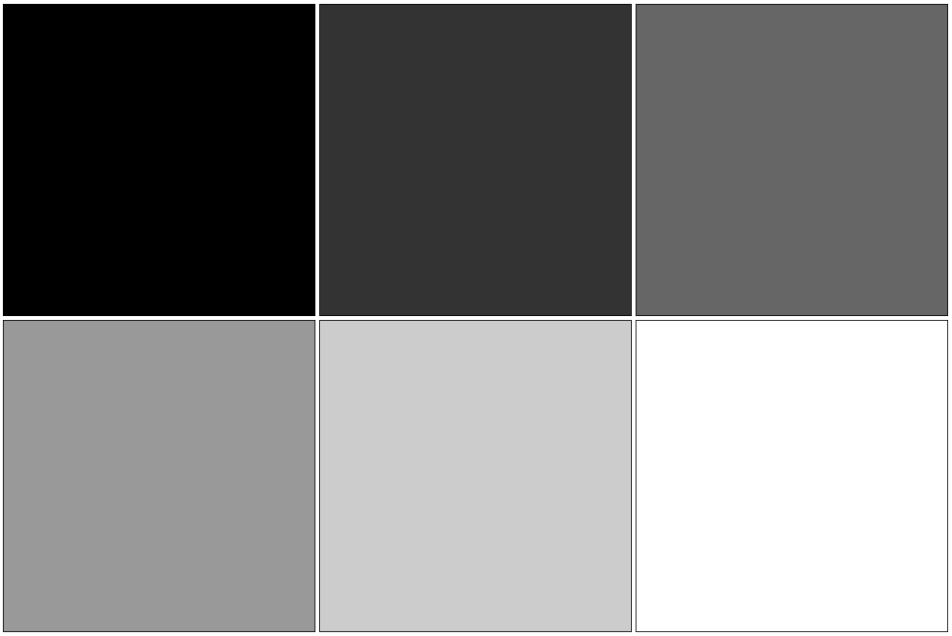


Figure E.19: Ground truth class map of the 6 texture images. Top row from left to right: T2-5, T2-4, T2-3; Bottom row from left to right: T2-2, T2-1, T2-0.

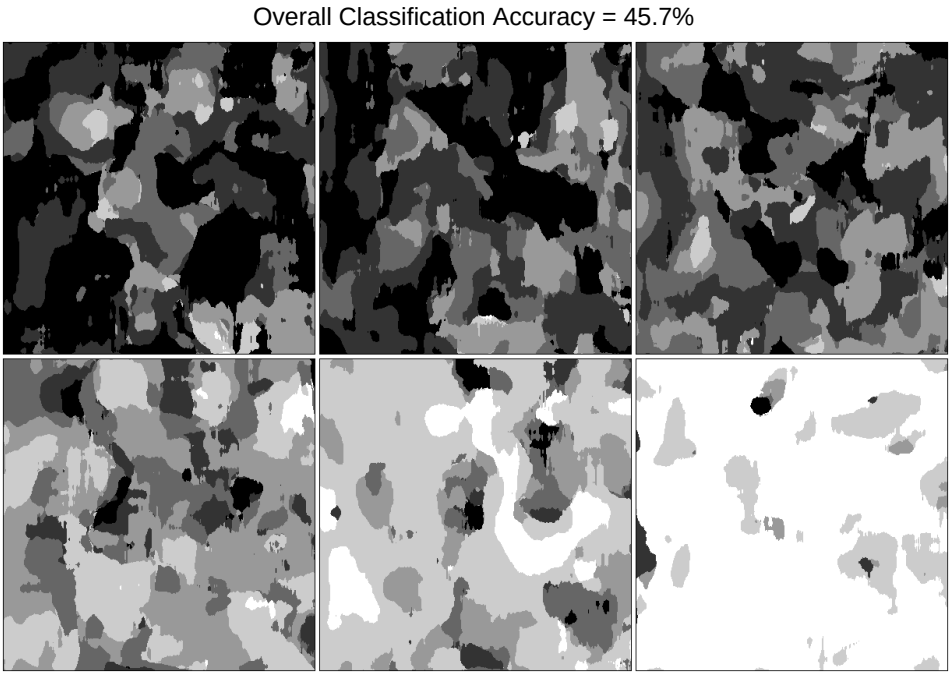


Figure E.20: Classification map using ICA (Exp. No. 2)

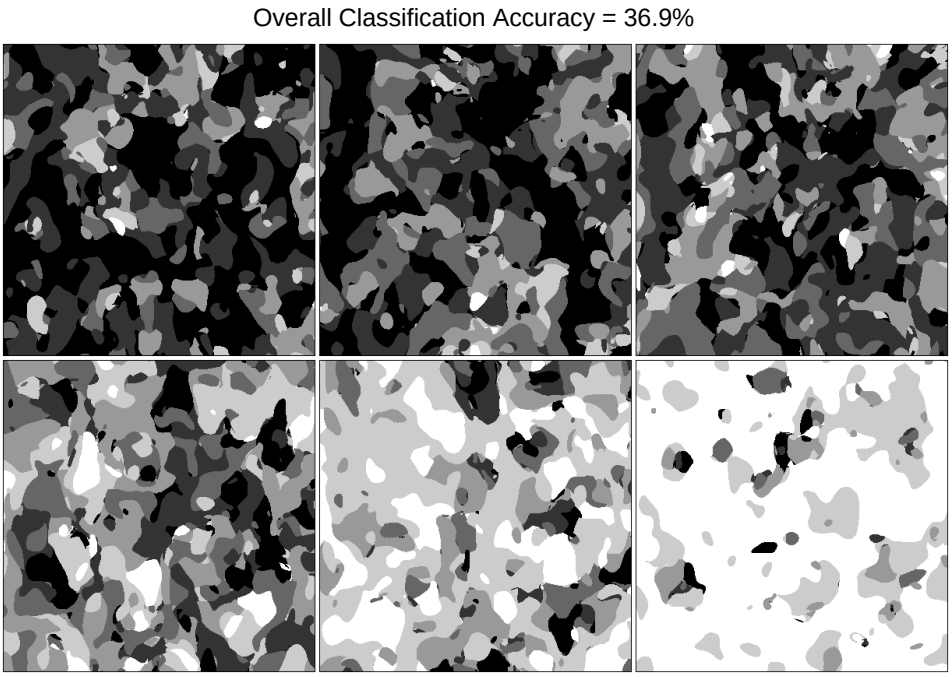


Figure E.21: Classification map using PCA (Exp. No. 2)

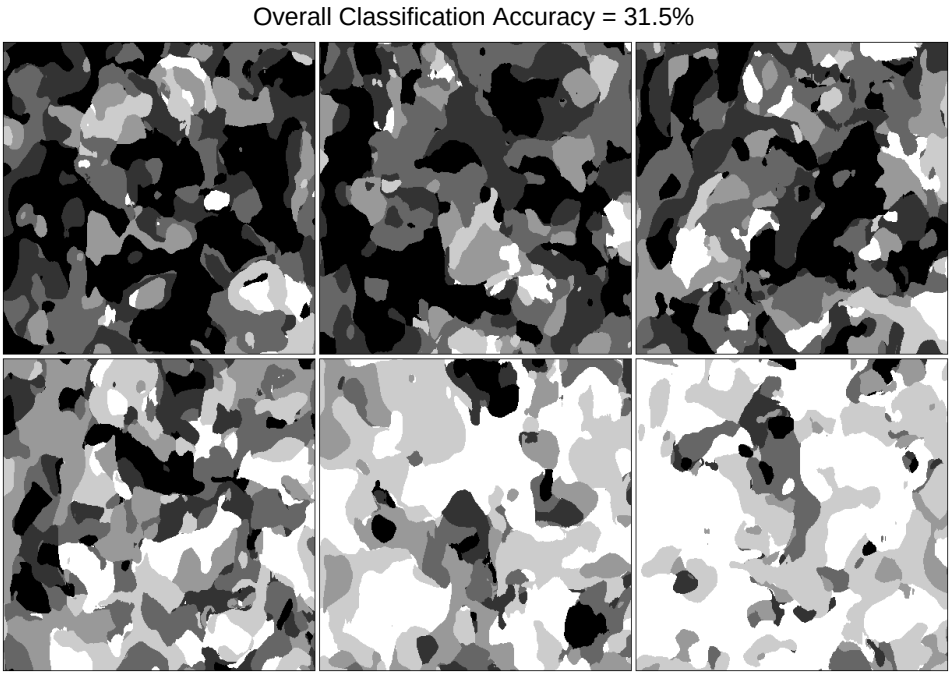


Figure E.22: Classification map using Gabor (Exp. No. 2)

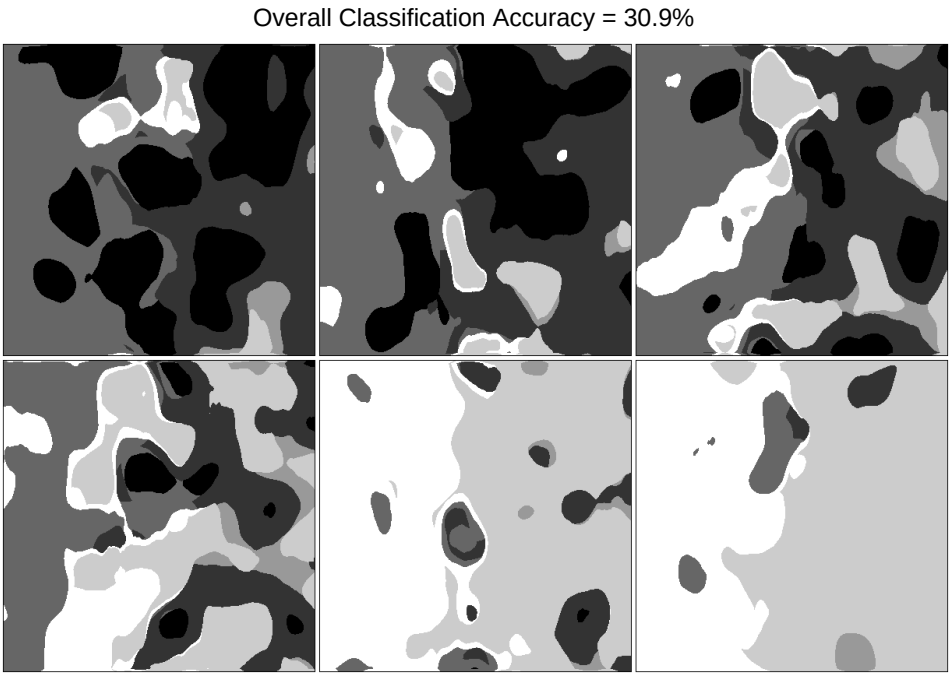


Figure E.23: Classification map using FWHM-ACVF (Exp. No. 2)

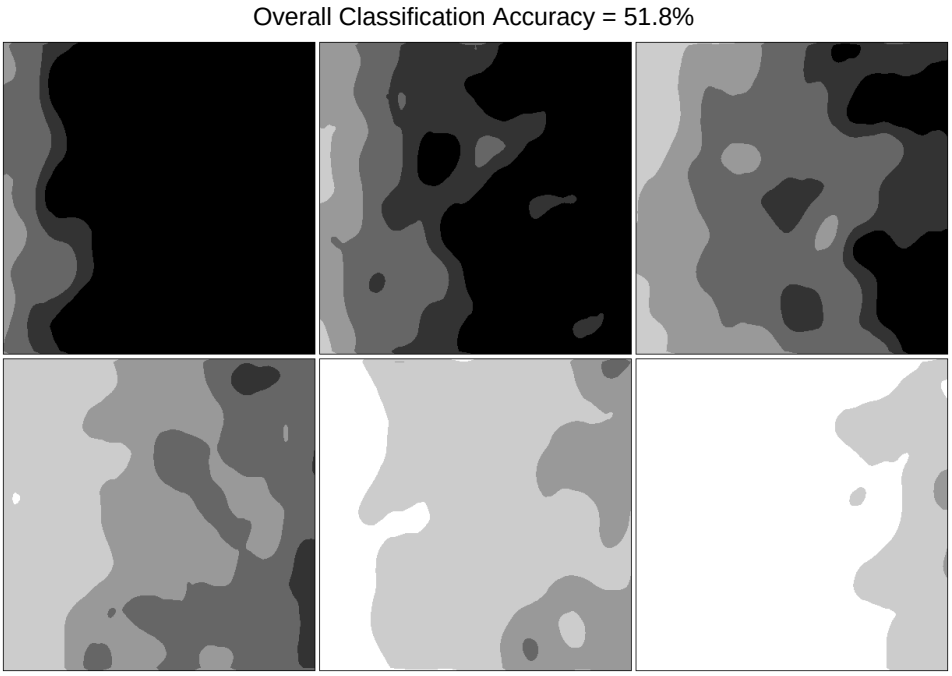


Figure E.24: Classification map using SNR-ALPHA (Exp. No. 2)

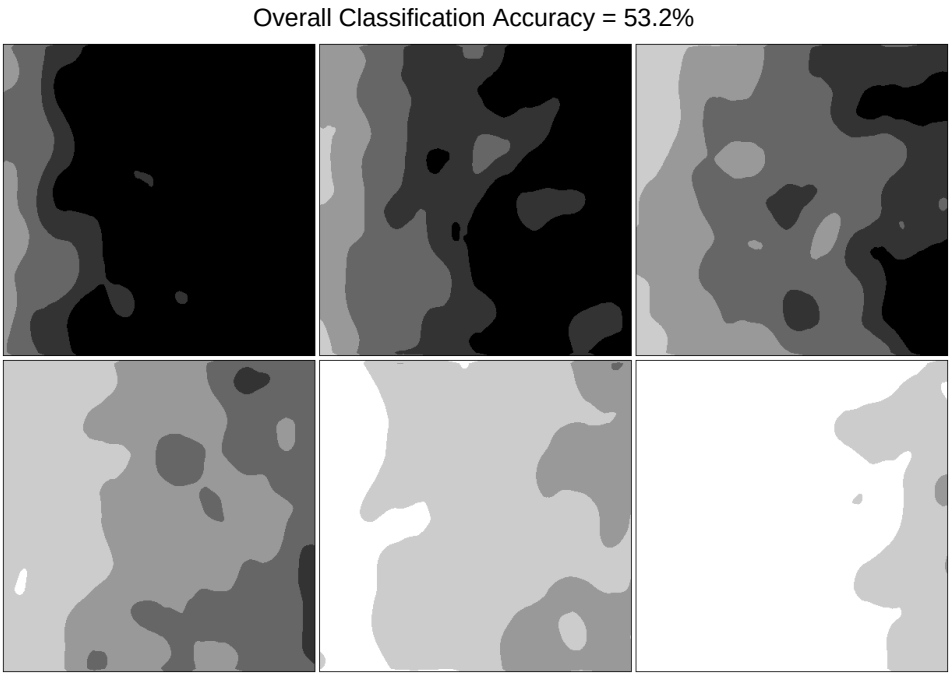
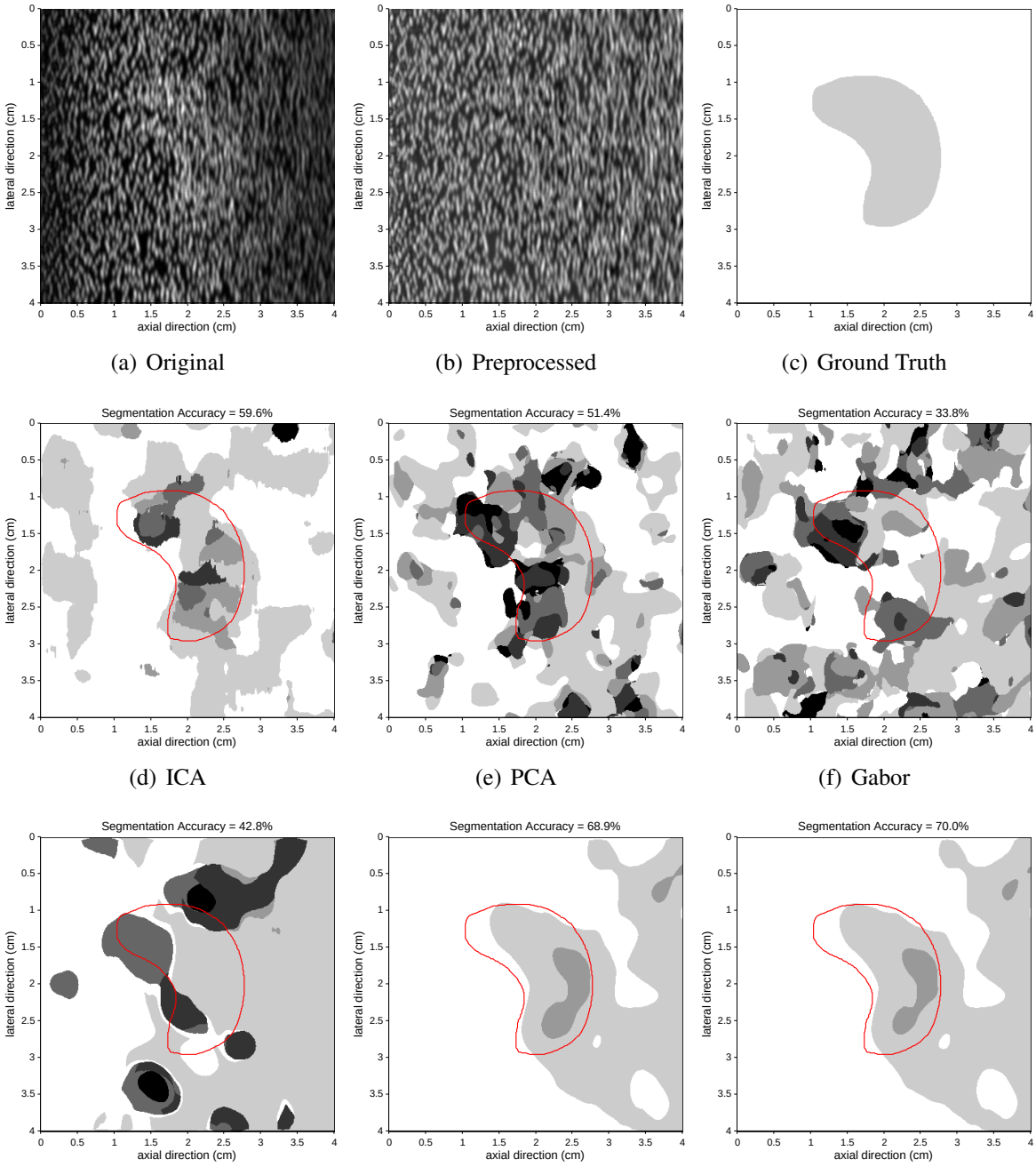


Figure E.25: Classification map using SNR-GAMMA (Exp. No. 2)

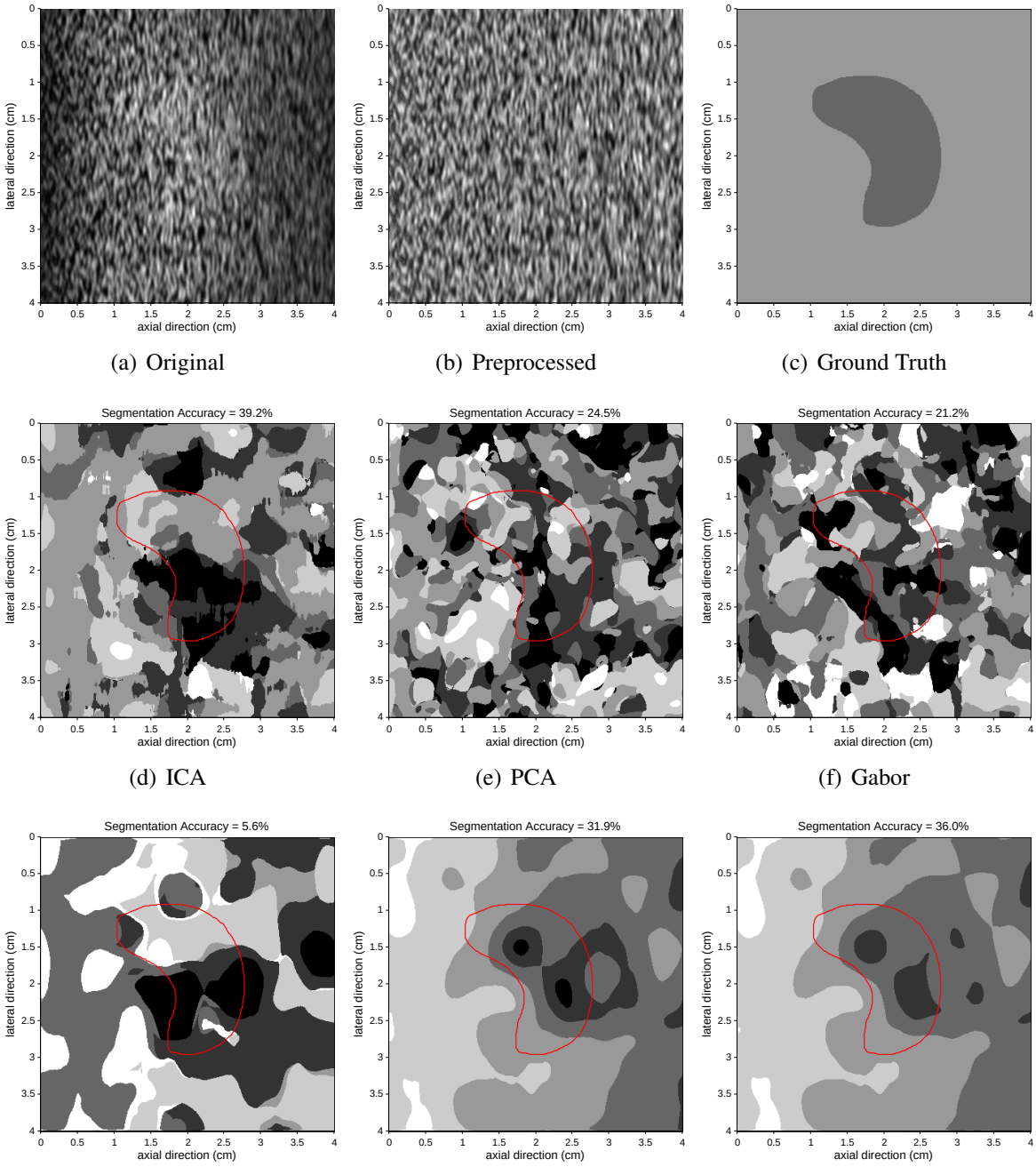


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.26: Two-component Texture Segmentation Results (Background:T2-0, Lesion:T2-1)

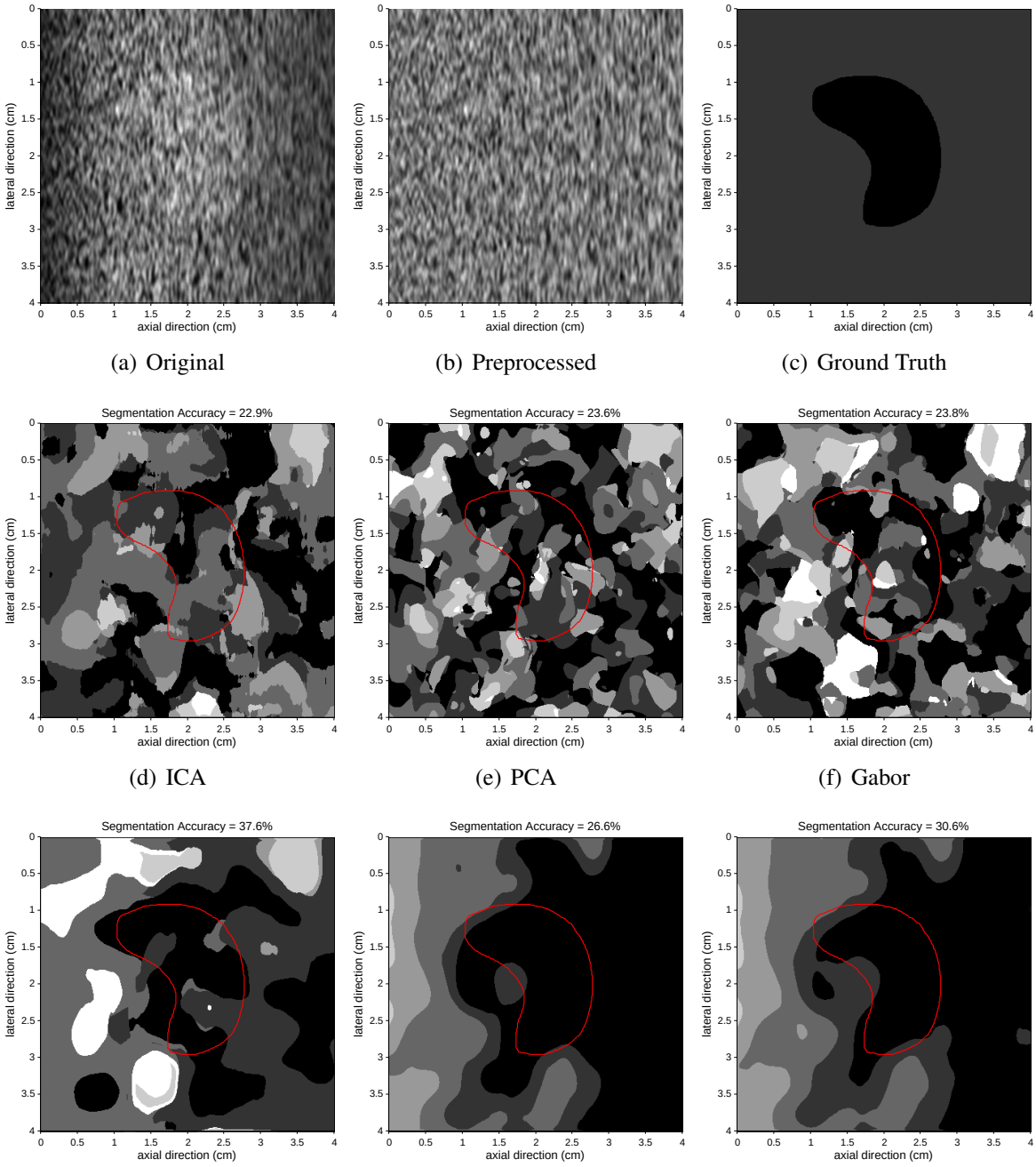


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.27: Two-component Texture Segmentation Results (Background:T2-2, Lesion:T2-3)

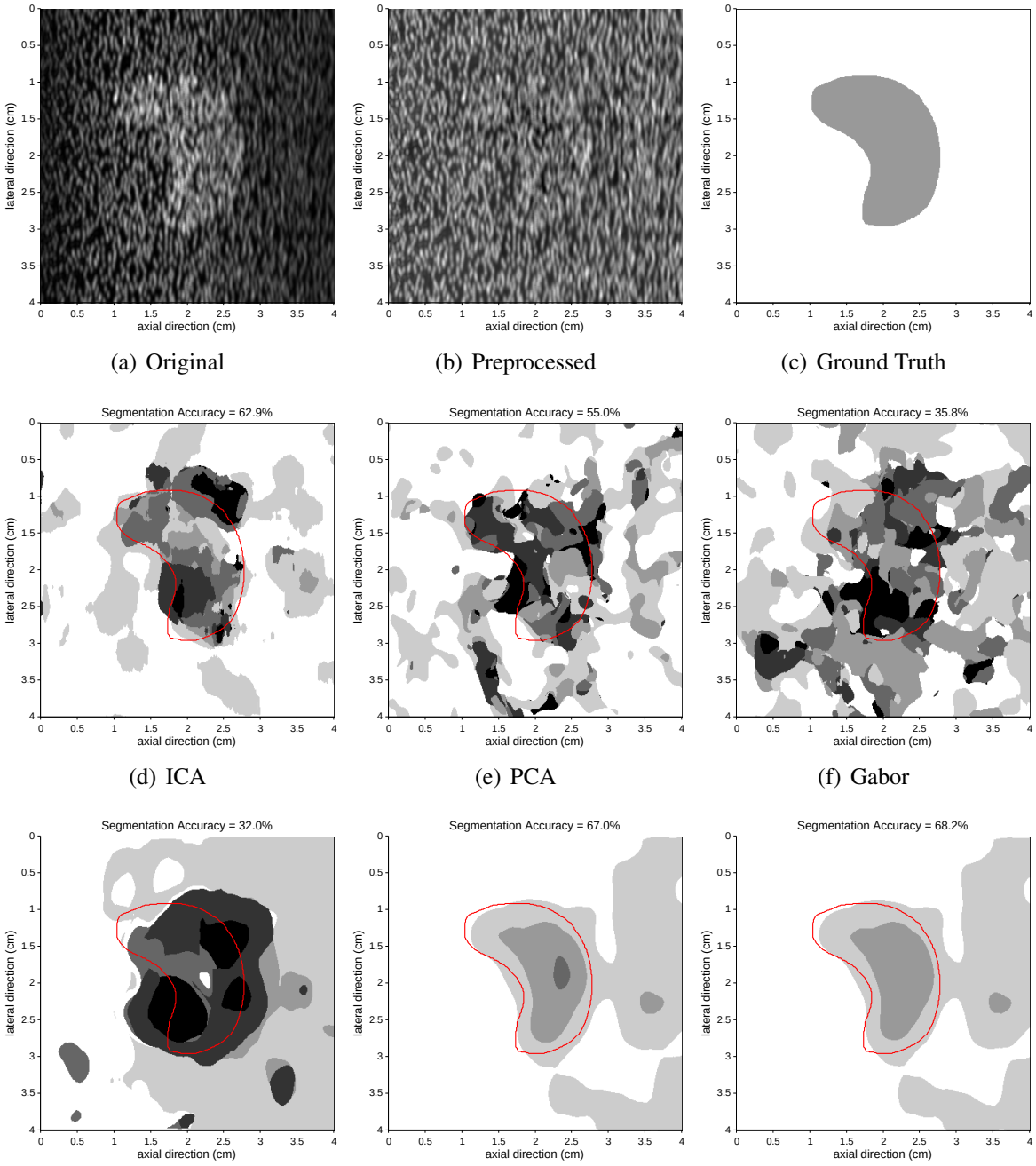


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.28: Two-component Texture Segmentation Results (Background:T2-4, Lesion:T2-5)

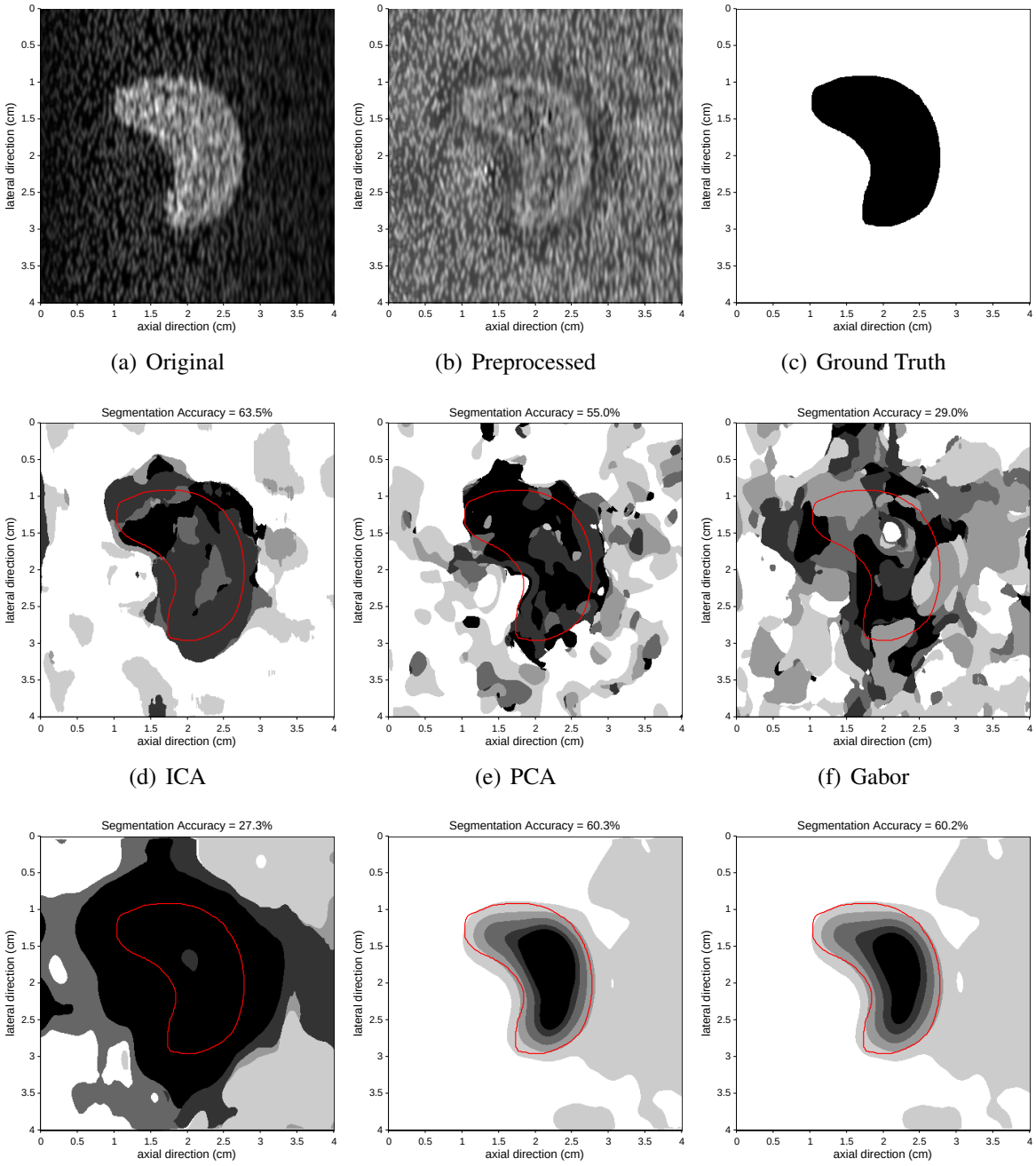


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.29: Two-component Texture Segmentation Results (Background:T2-0, Lesion:T2-2)



(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.30: Two-component Texture Segmentation Results (Background:T2-0, Lesion:T2-5)

E.3 Results of Exp. No. 3

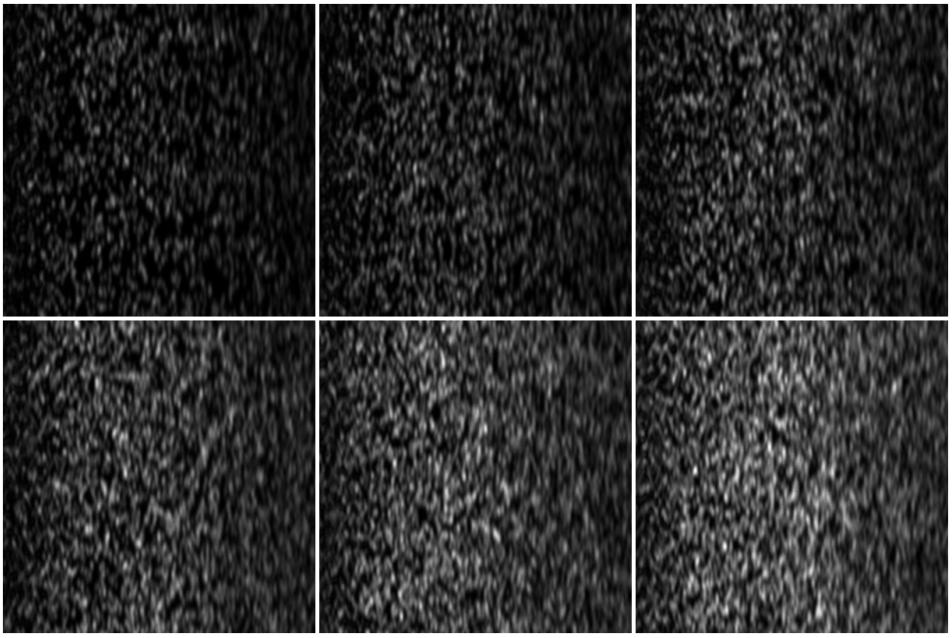


Figure E.31: 6 original B-mode texture images for classification. Top row from left to right: T3-0, T3-1, T3-2; Bottom row from left to right: T3-3, T3-4, T3-5.

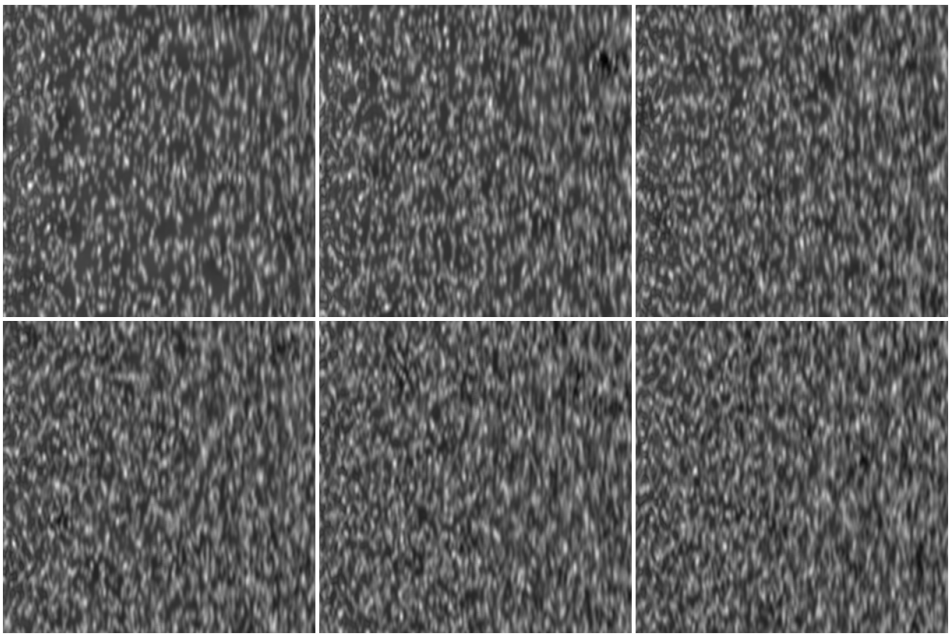
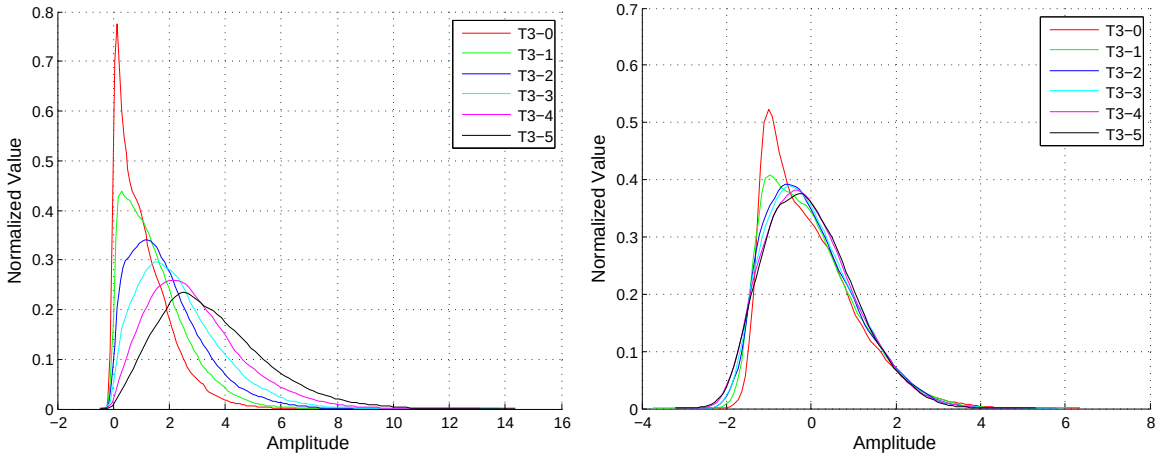


Figure E.32: 6 preprocessed B-mode texture images for classification. Top row from left to right: T3-0, T3-1, T3-2; Bottom row from left to right: T3-3, T3-4, T3-5.



(a) Before preprocessing (b) After preprocessing

Figure E.33: Histogram of the 6 texture images before and after preprocessing

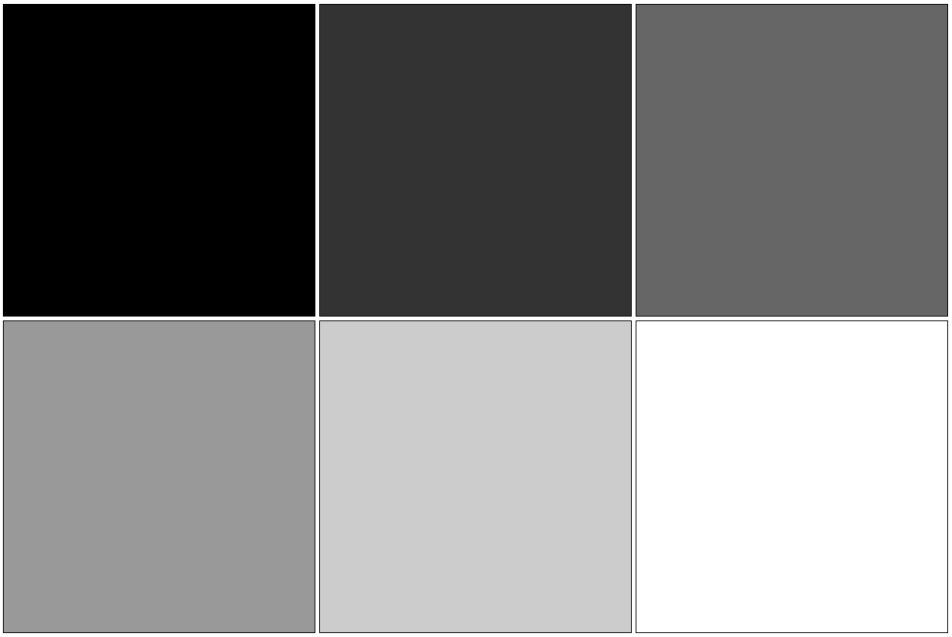


Figure E.34: Ground truth class map of the 6 texture images.

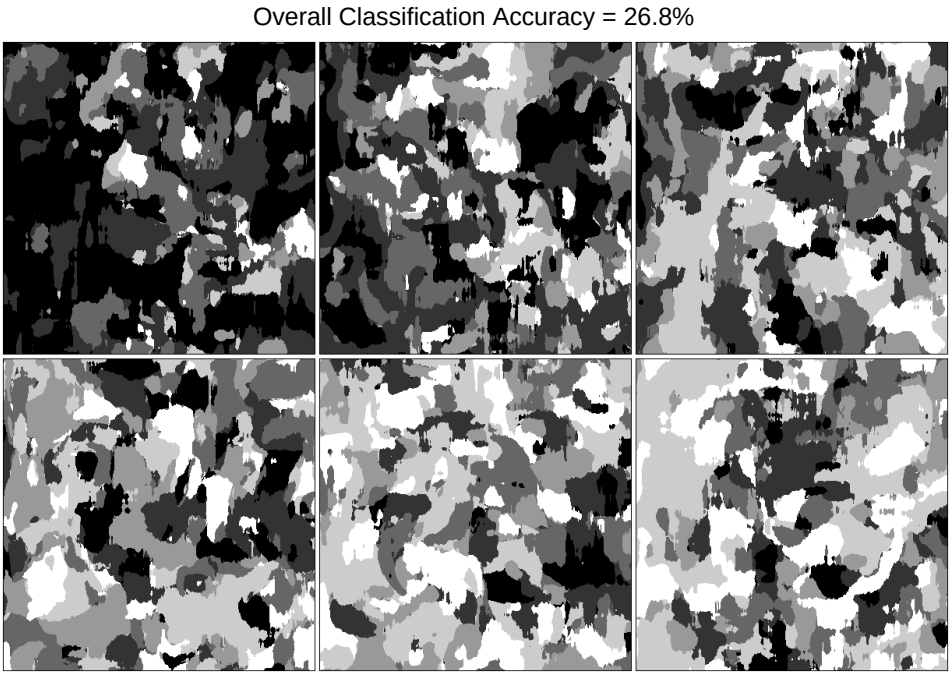


Figure E.35: Classification map using ICA (Exp. No. 3)

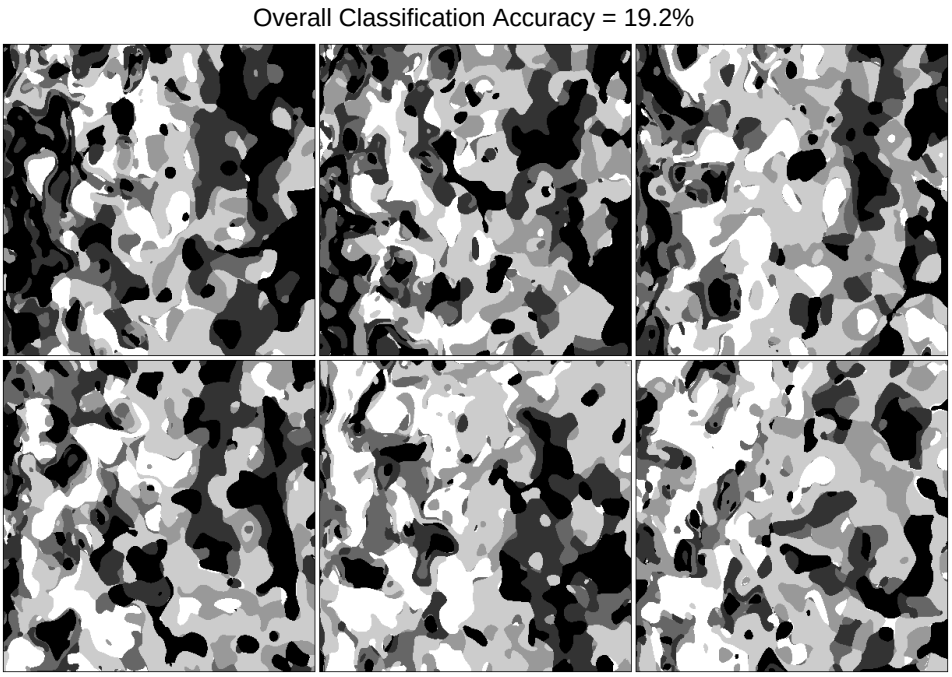


Figure E.36: Classification map using PCA (Exp. No. 3)

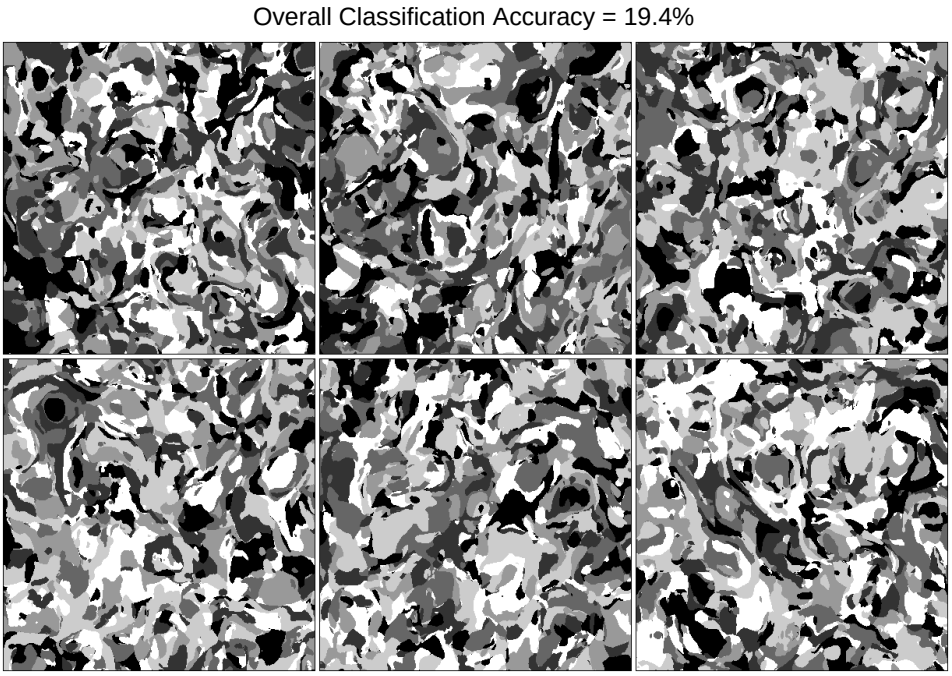


Figure E.37: Classification map using Gabor (Exp. No. 3)

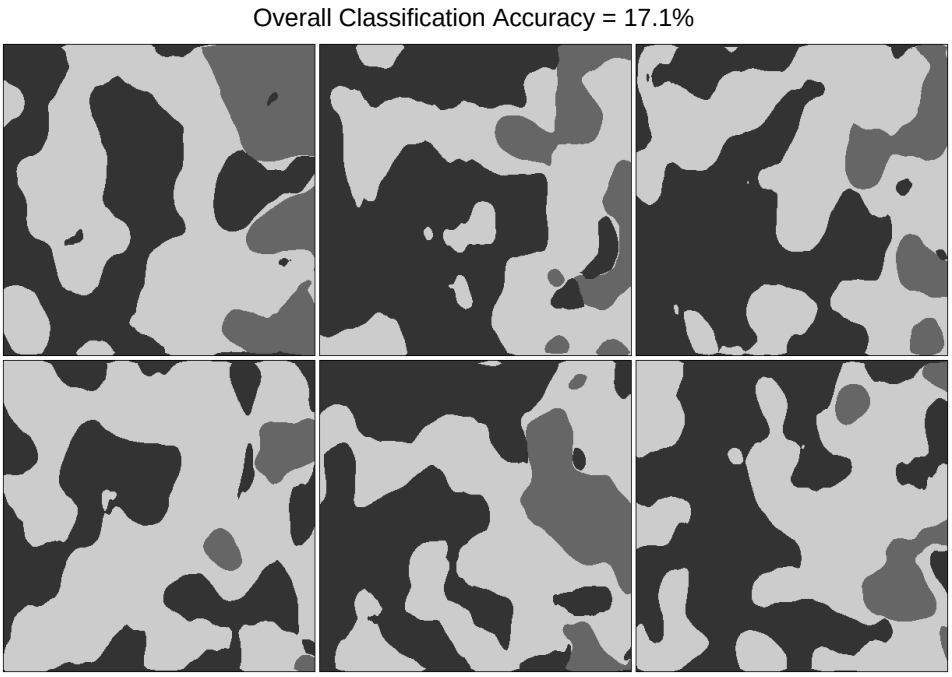


Figure E.38: Classification map using FWHM-ACVF (Exp. No. 3)

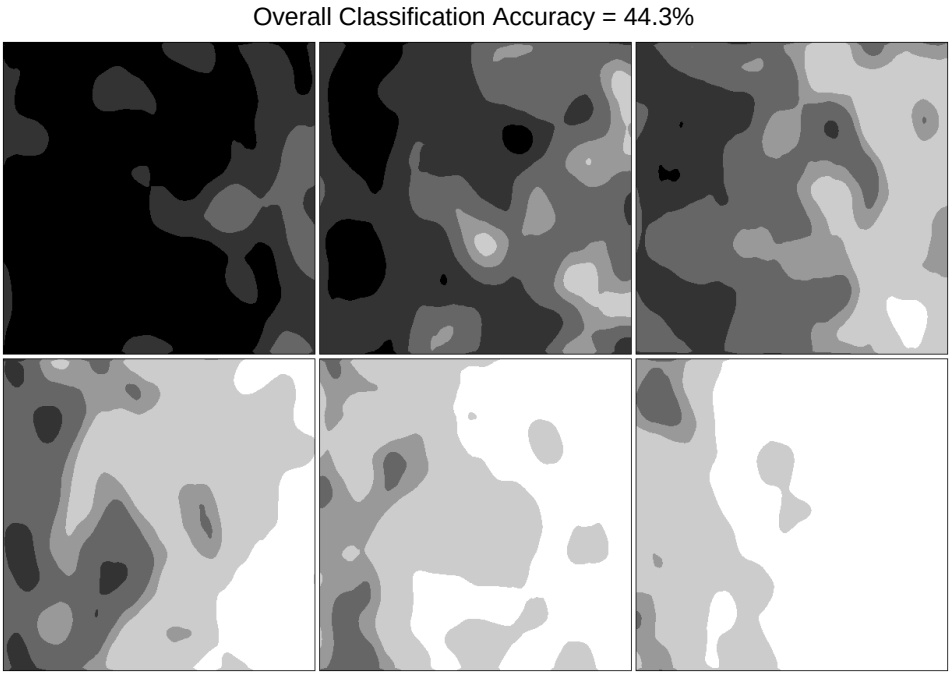


Figure E.39: Classification map using SNR-ALPHA (Exp. No. 3)

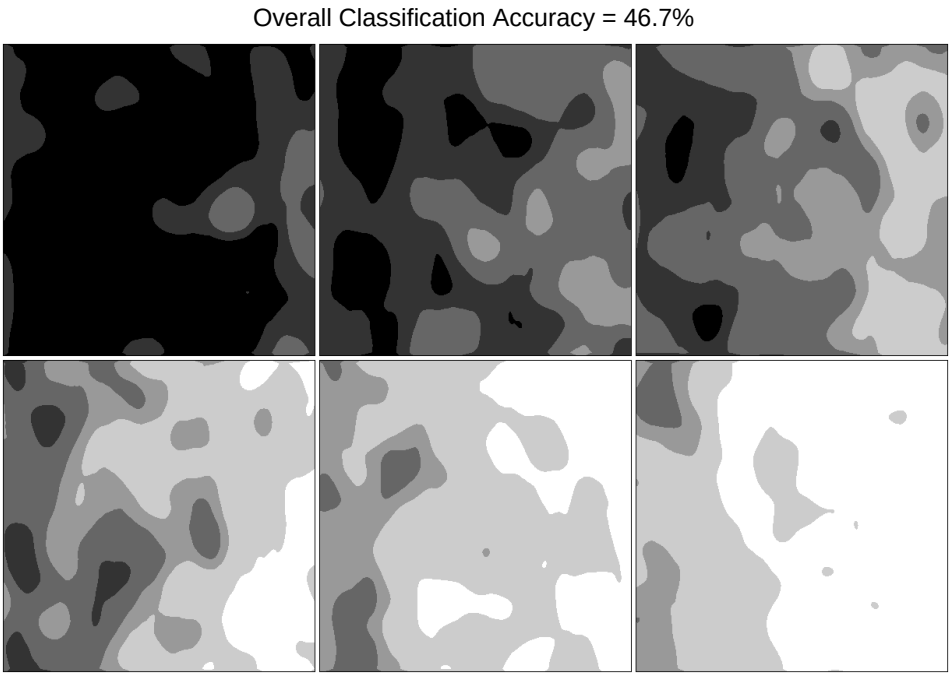


Figure E.40: Classification map using SNR-GAMMA (Exp. No. 3)

E.4 Results of Exp. No. 4

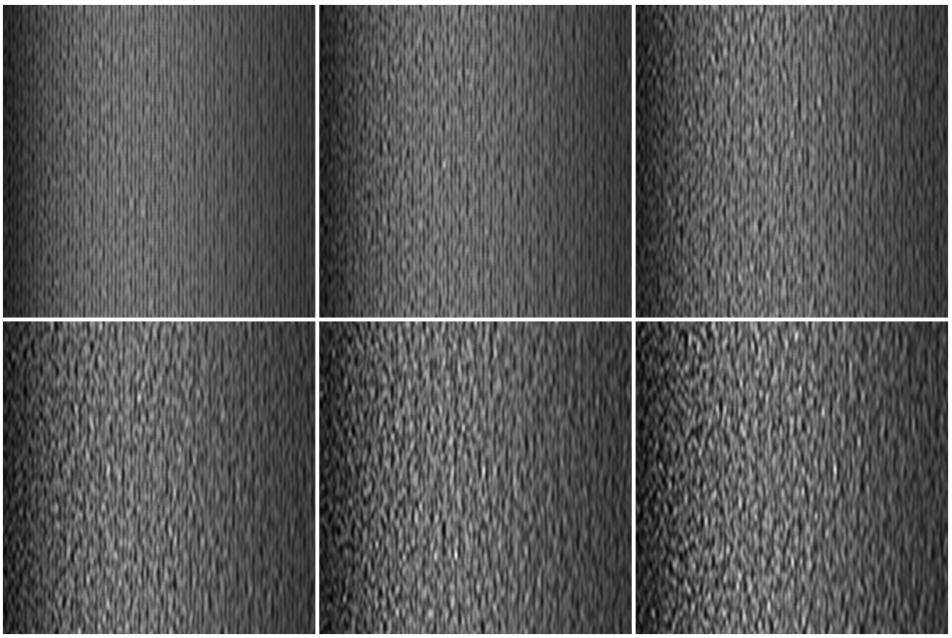


Figure E.41: 6 original texture images for classification.

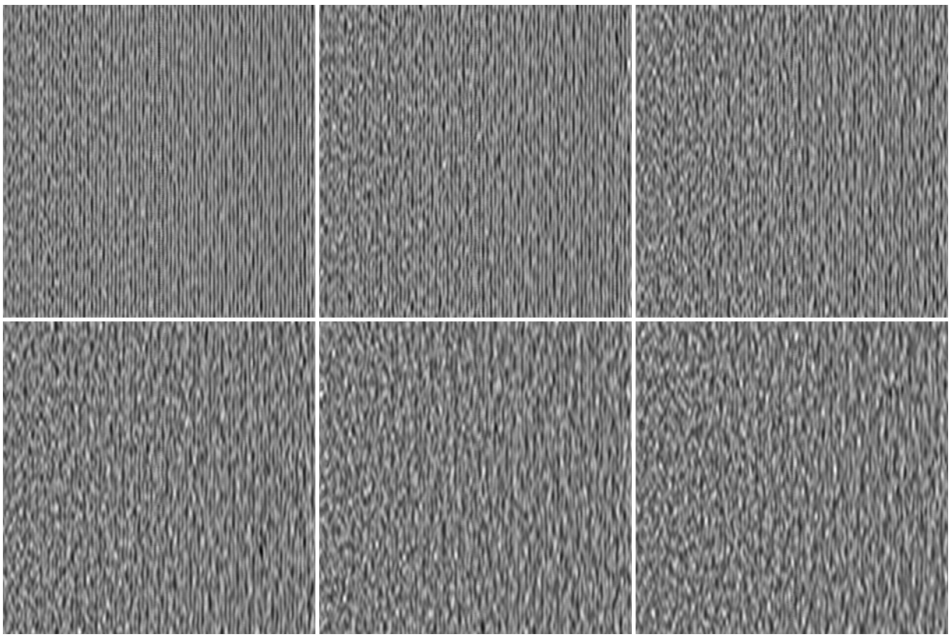
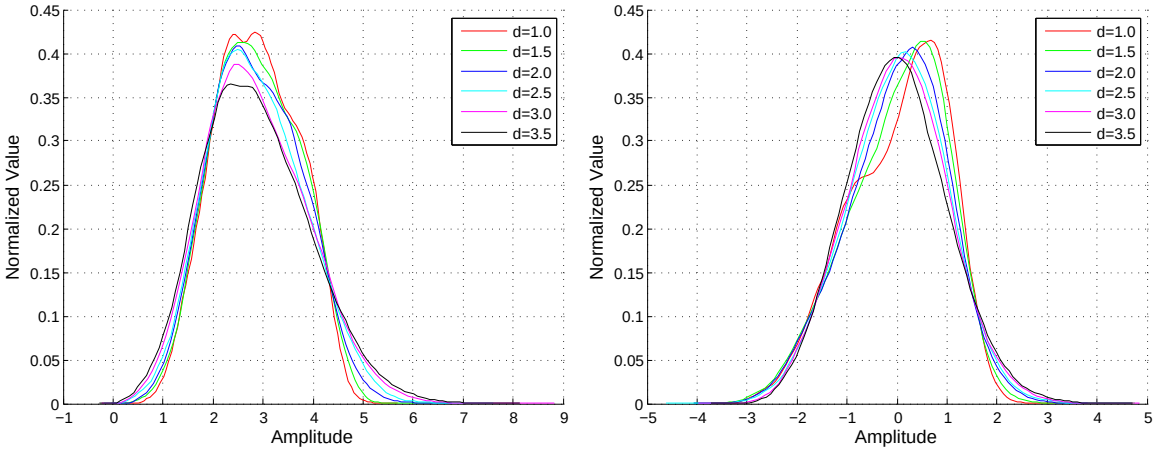


Figure E.42: 6 preprocessed envelope texture images for classification.



(a) Before preprocessing (b) After preprocessing

Figure E.43: Histogram of the 6 texture images before and after preprocessing

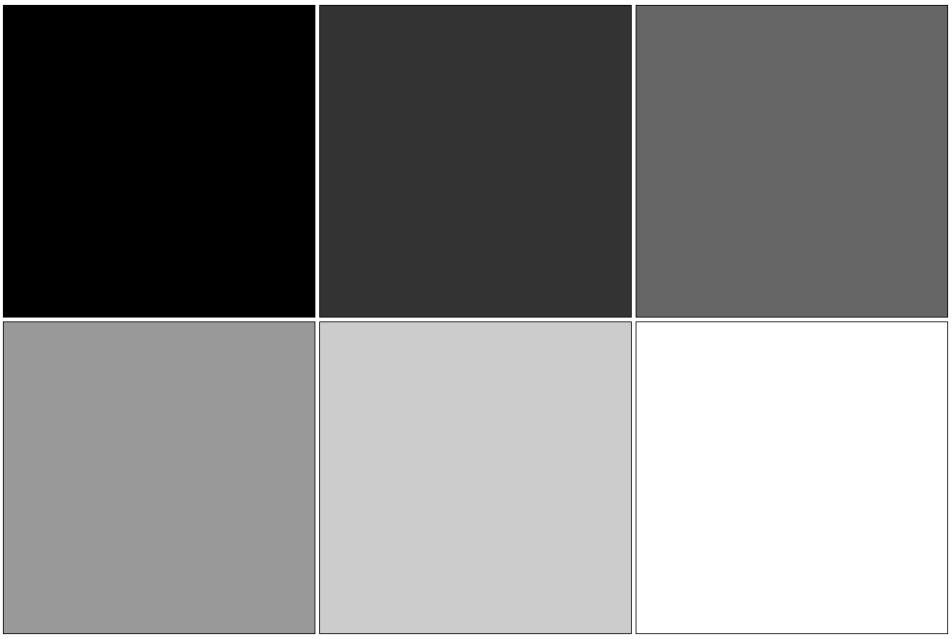


Figure E.44: Ground truth class map of the 6 texture images.

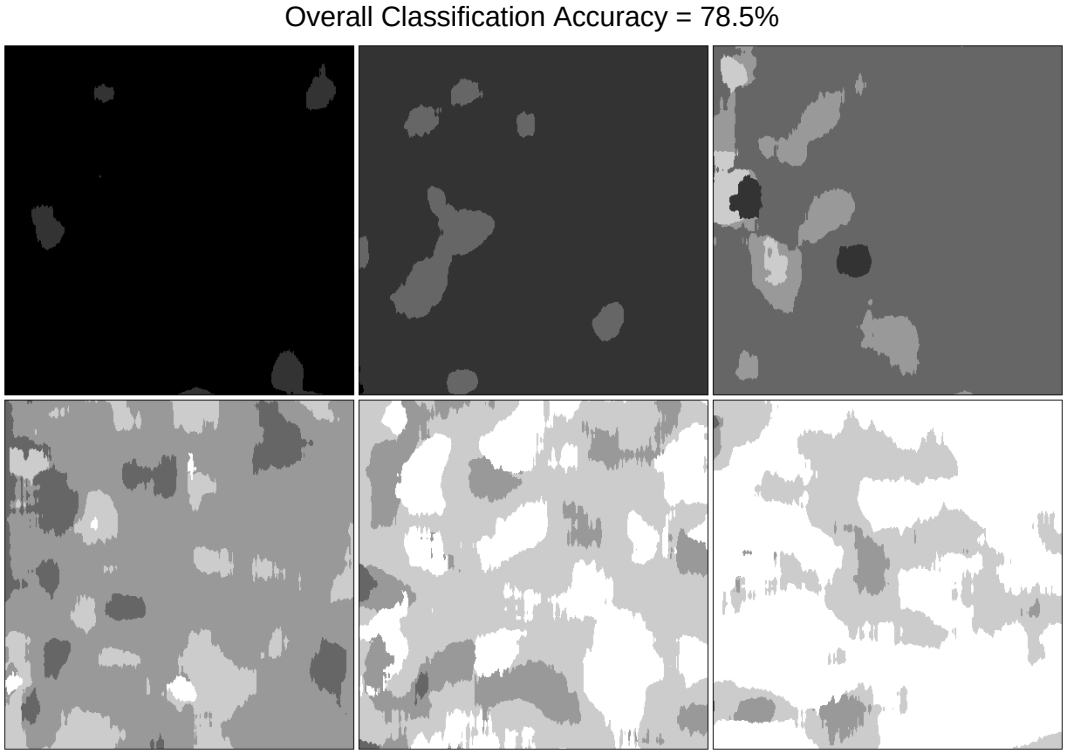


Figure E.45: Classification result using ICA (Exp. No. 4)

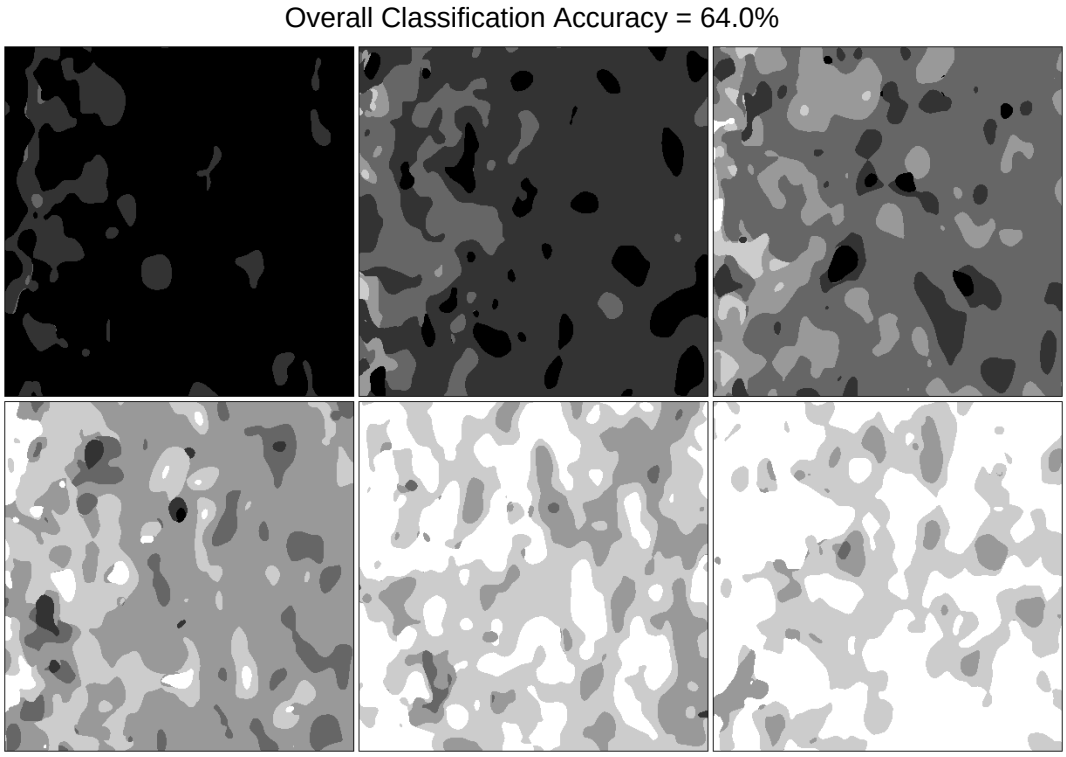


Figure E.46: Classification result using PCA (Exp. No. 4)

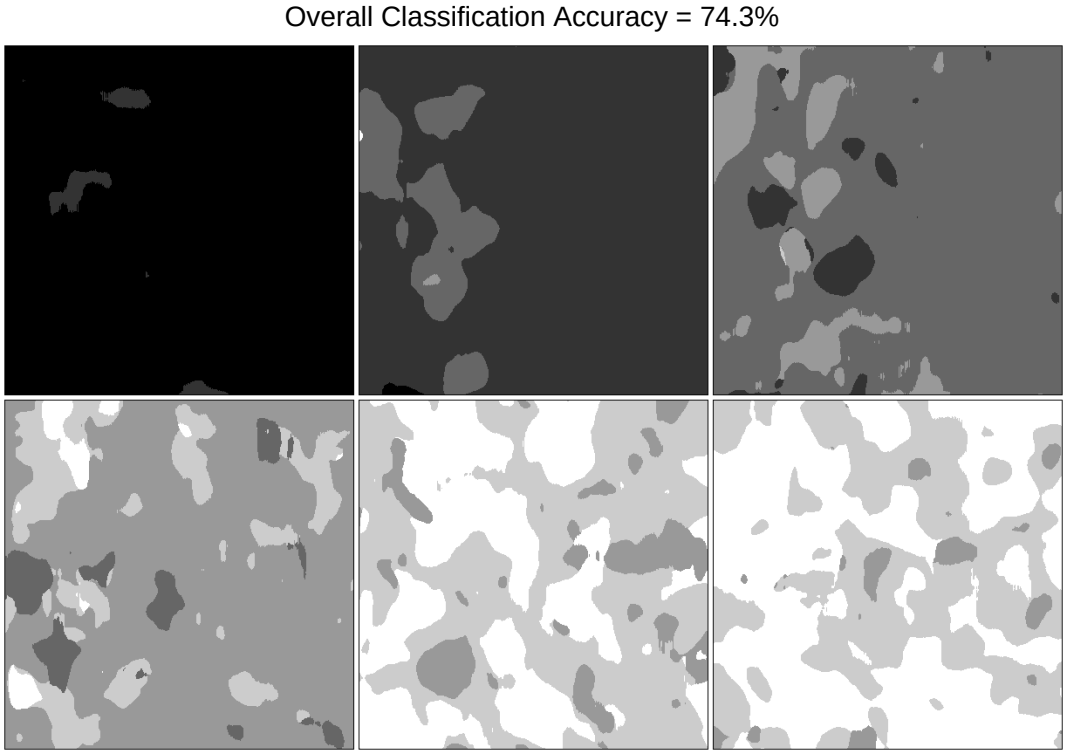


Figure E.47: Classification result using Gabor (Exp. No. 4)

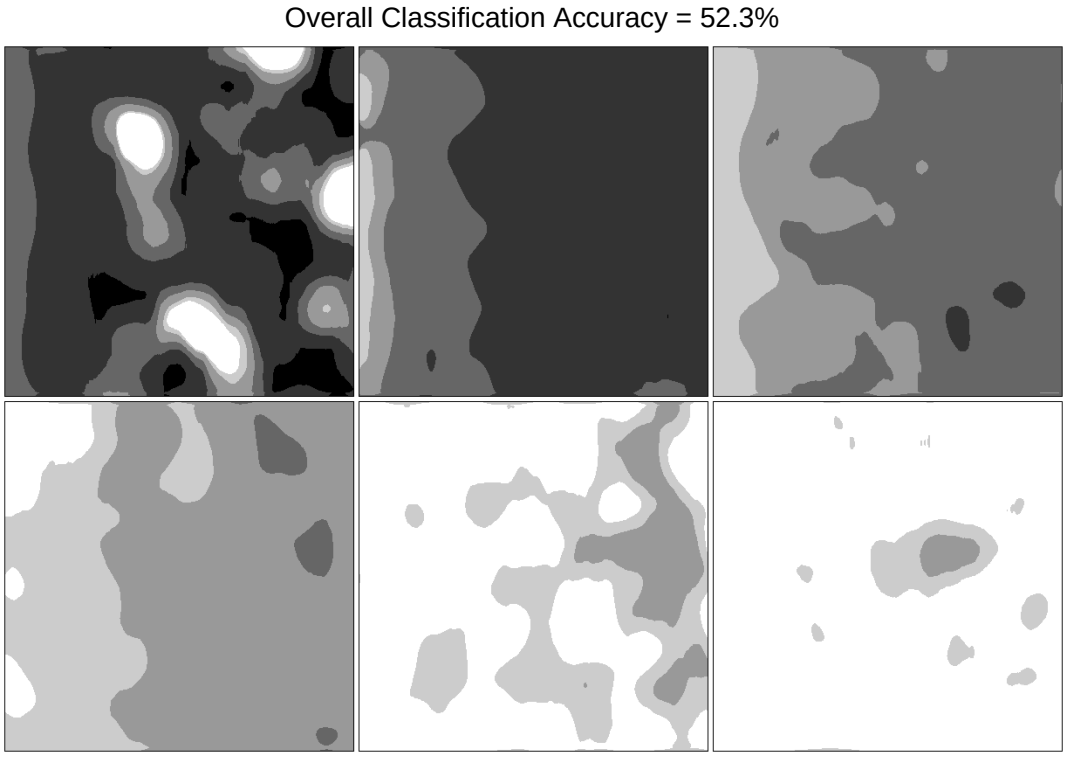


Figure E.48: Classification result using FWHM-ACVF (Exp. No. 4)

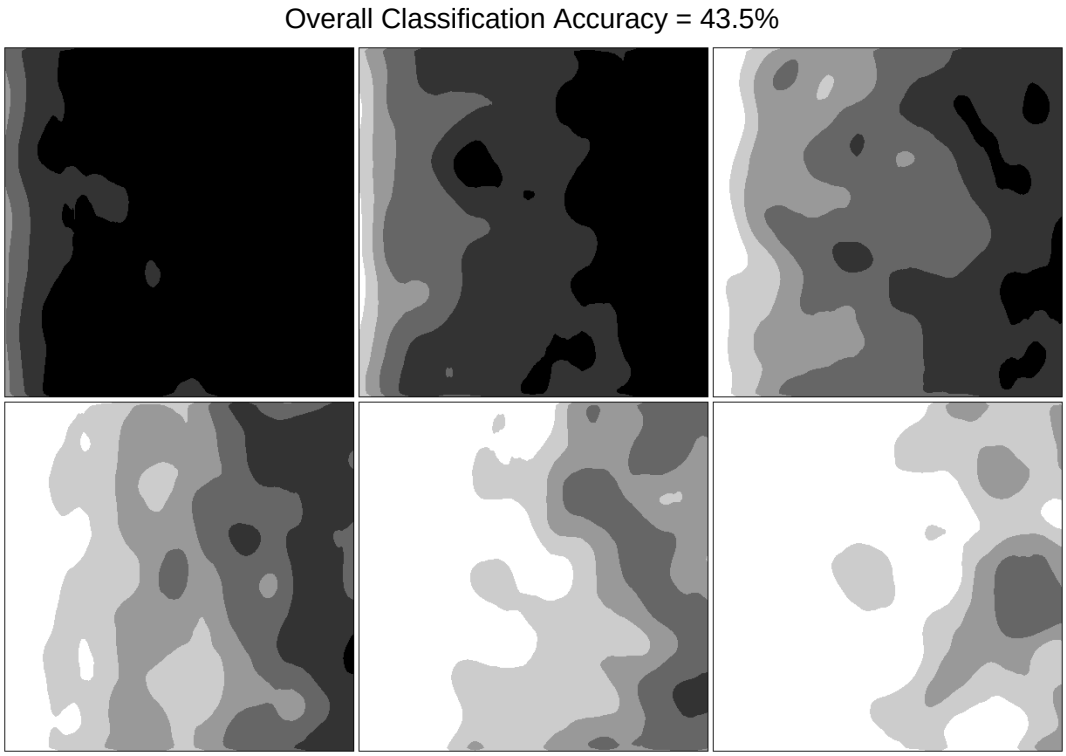


Figure E.49: Classification result using SNR-ALPHA (Exp. No. 4)

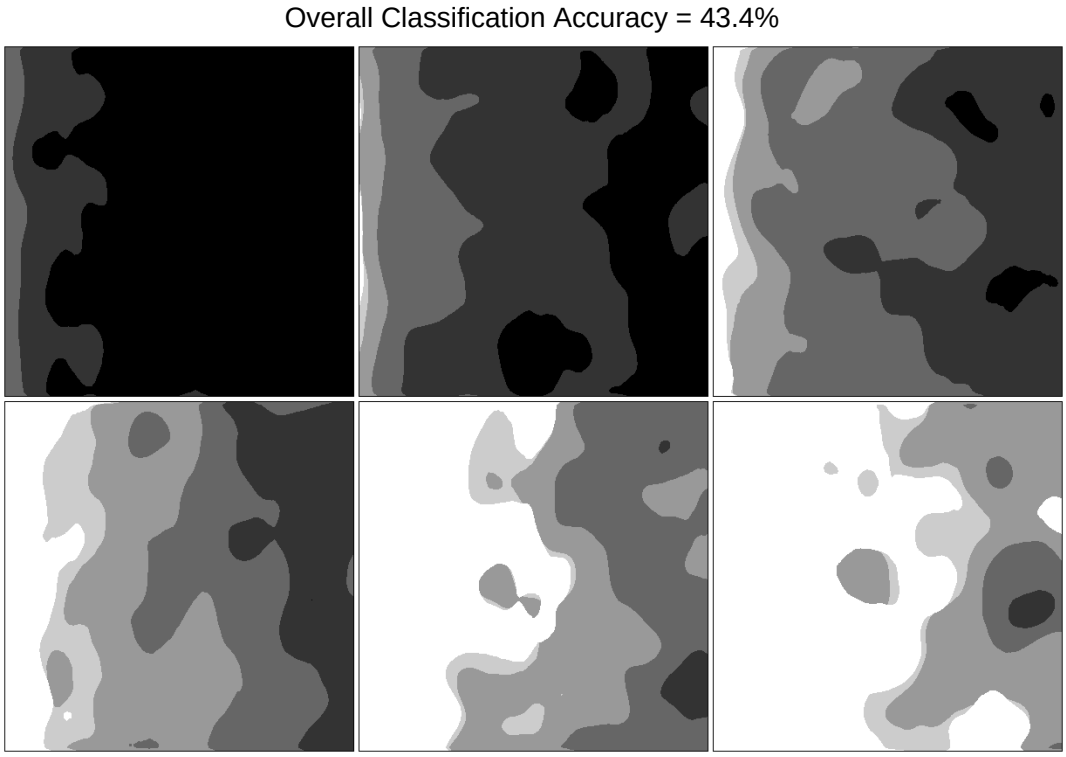
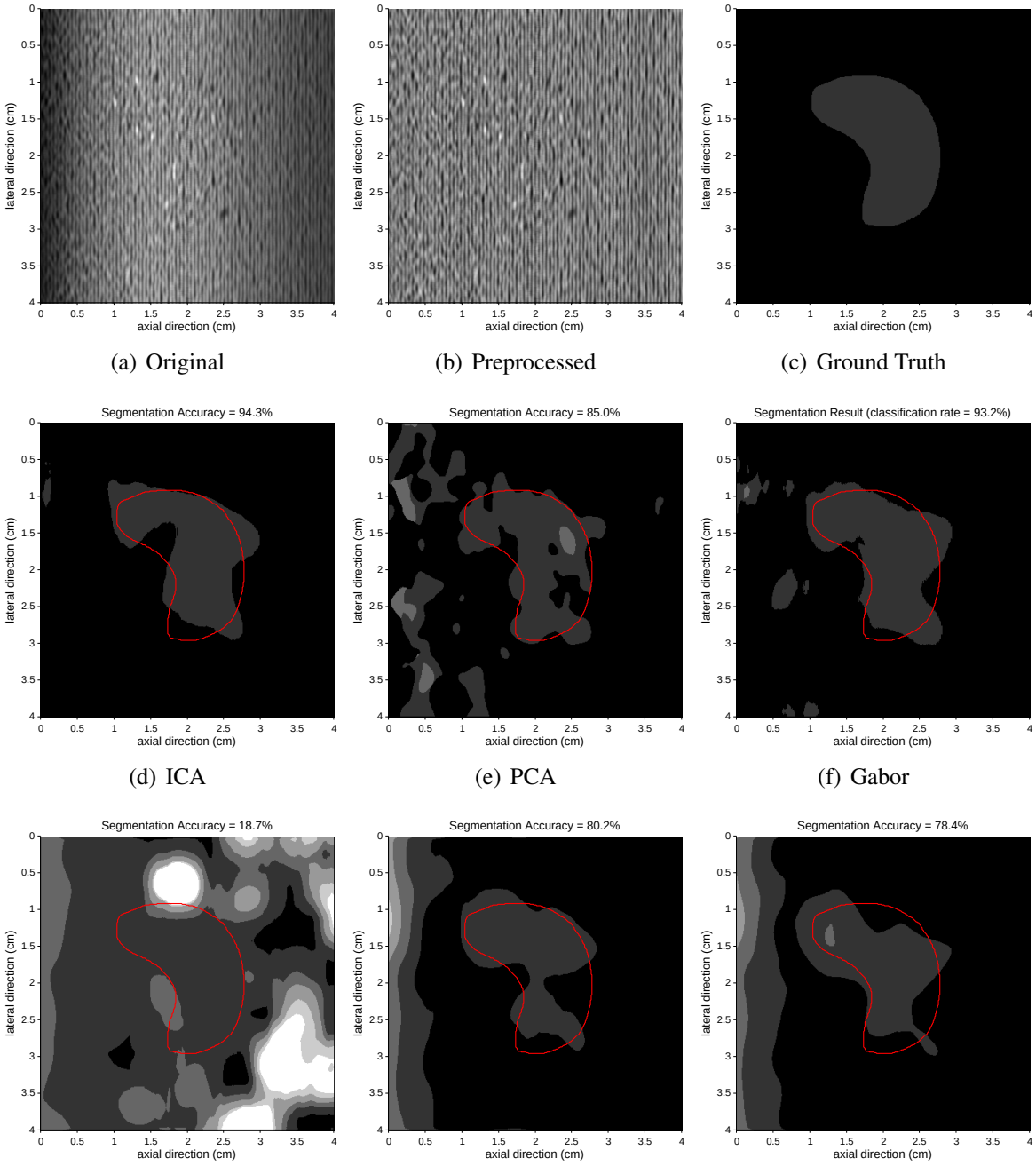


Figure E.50: Classification result using SNR-GAMMA (Exp. No. 4)

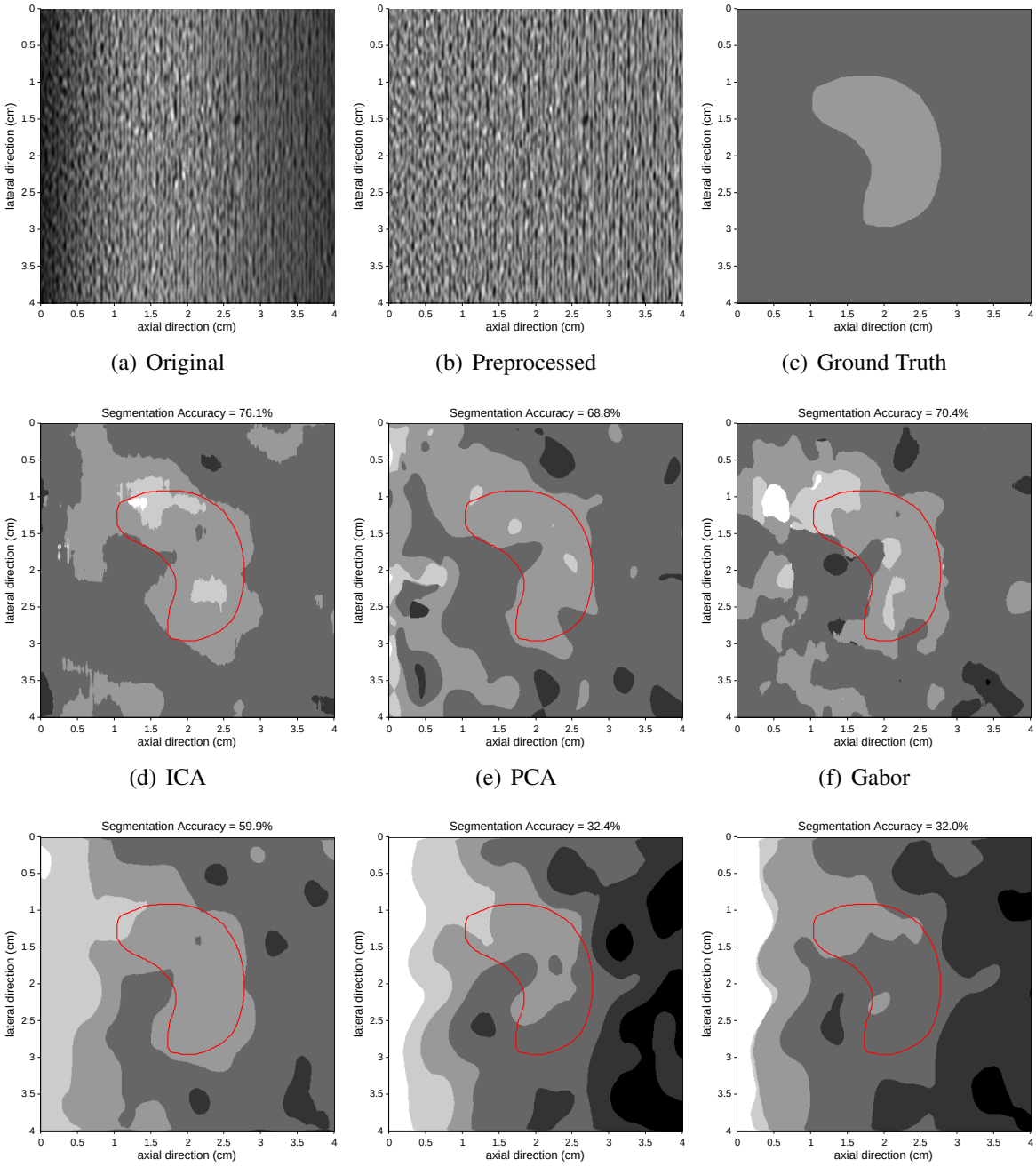


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.51: Two-component Texture Segmentation Results (Background:T4-0, Lesion:T4-1)

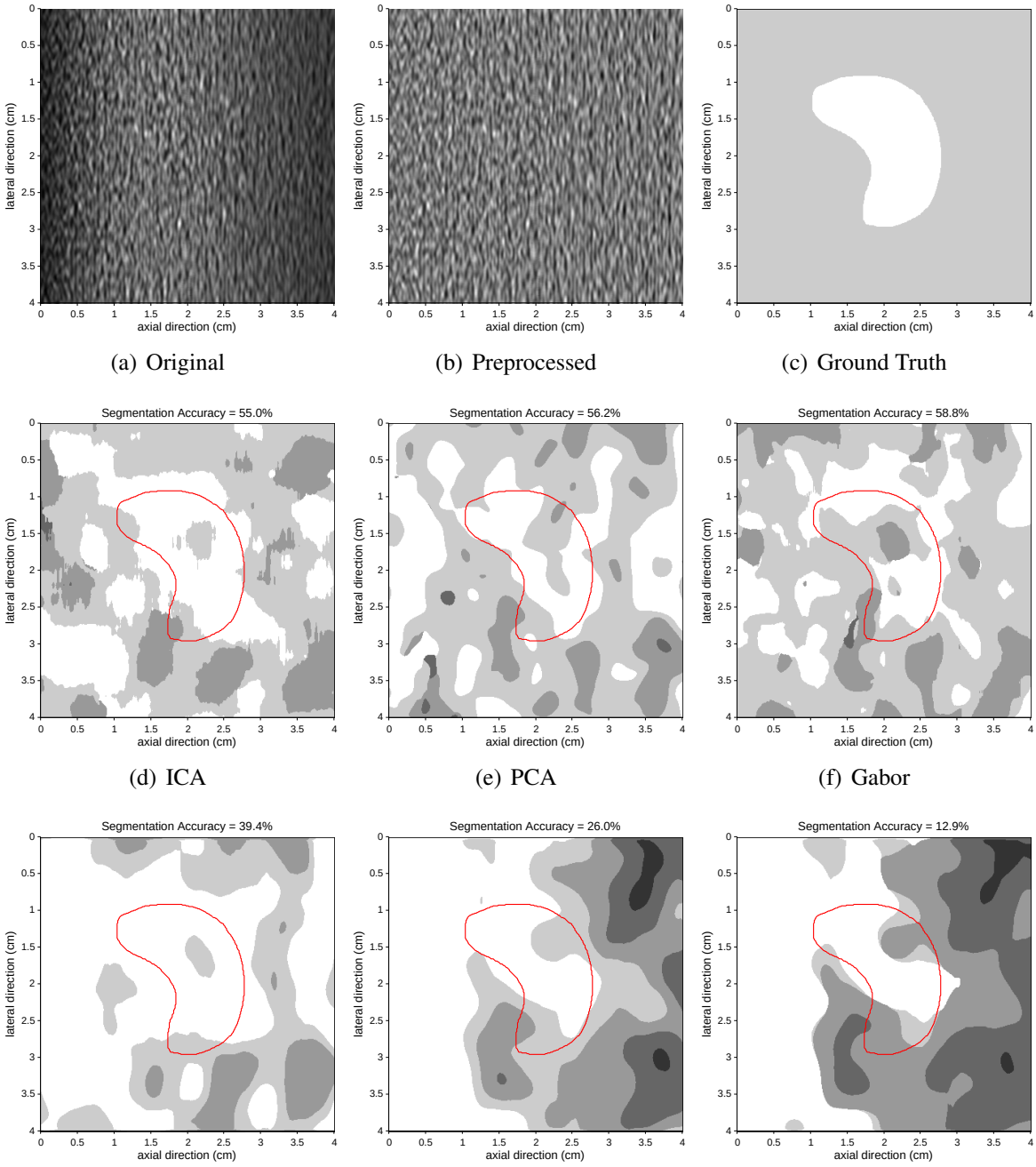


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.52: Two-component Texture Segmentation Results (Background:T4-2, Lesion:T4-3)

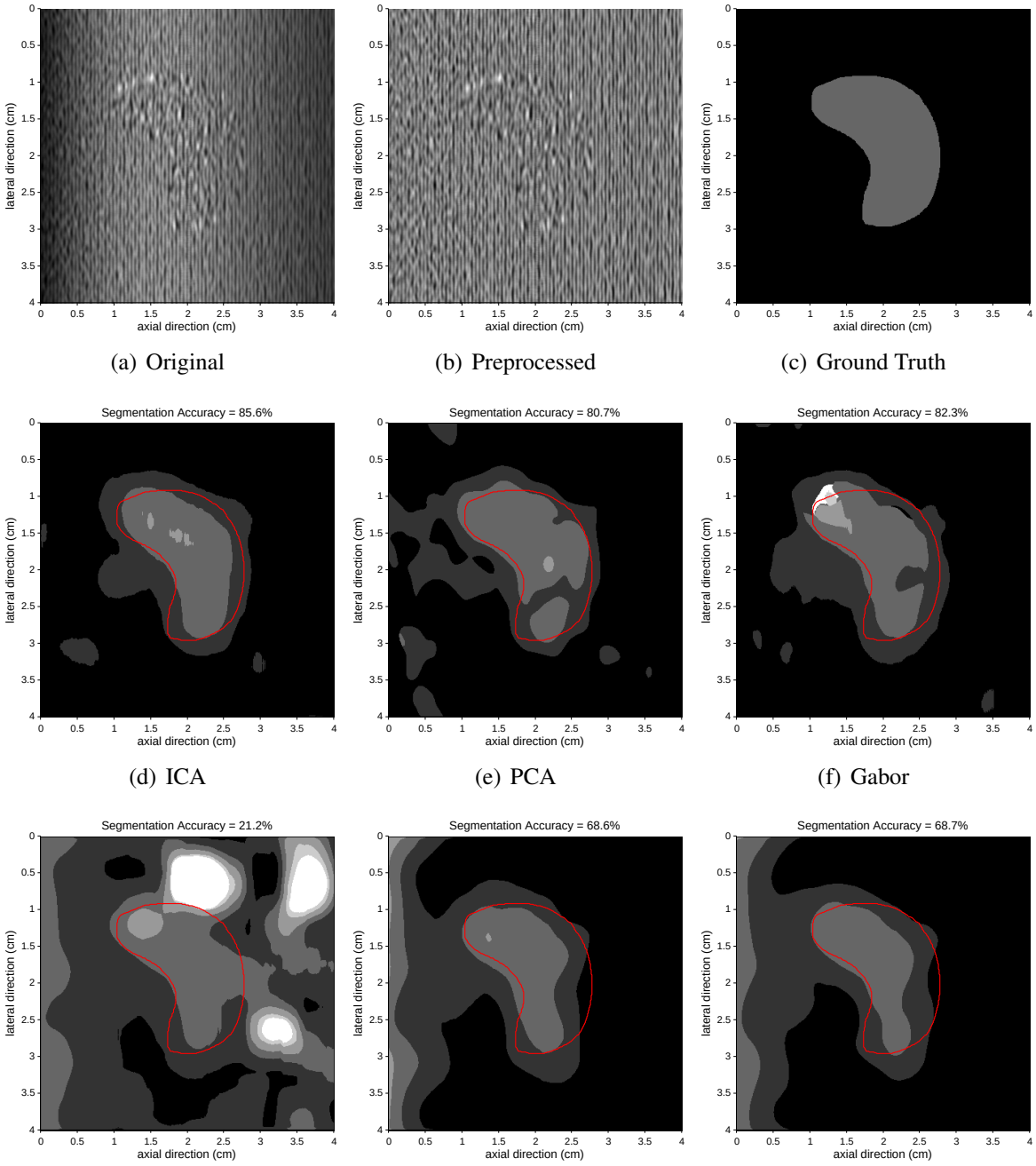


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.53: Two-component Texture Segmentation Results (Background:T4-4, Lesion:T4-5)

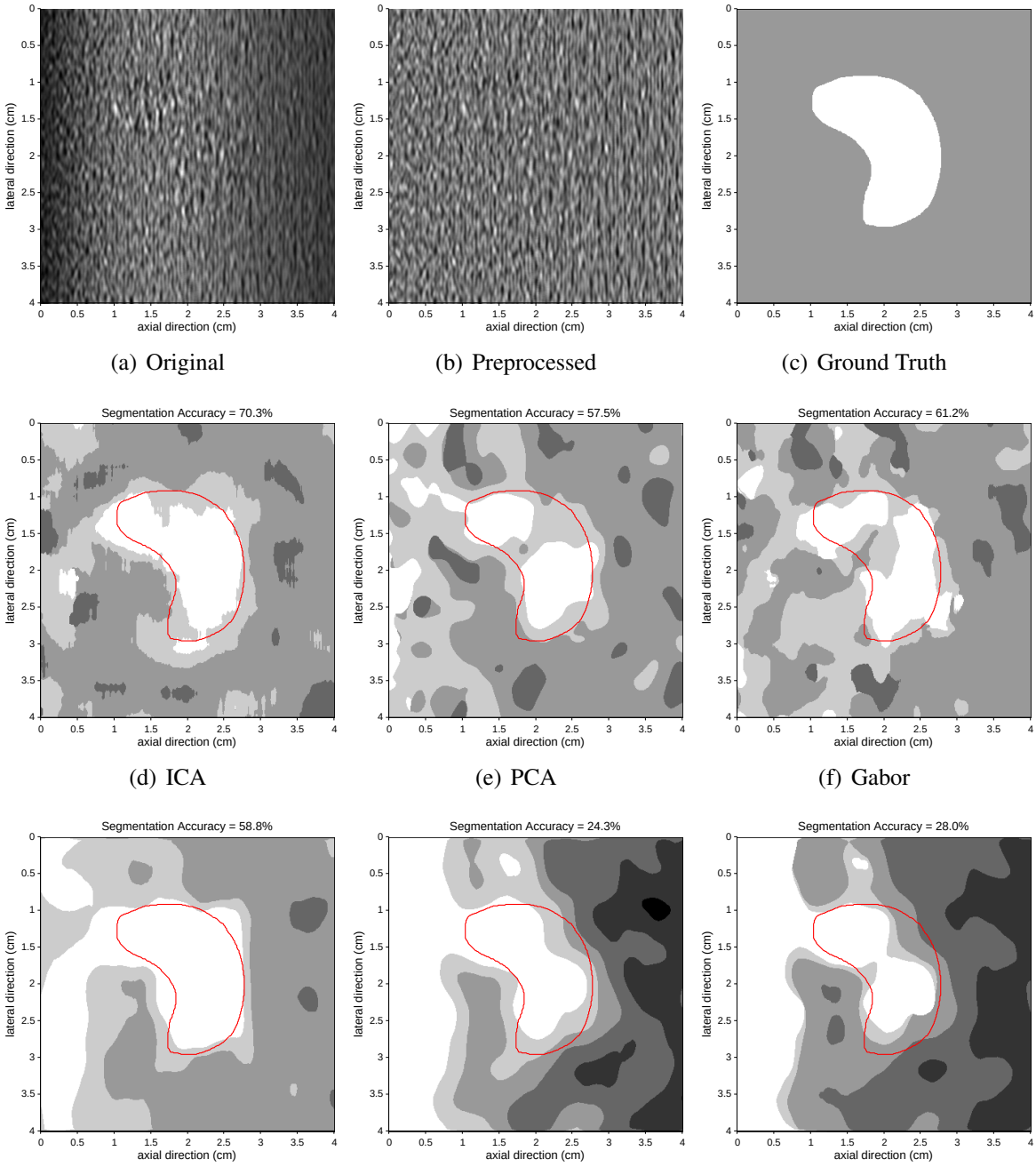


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.54: Two-component Texture Segmentation Results (Background:T4-0, Lesion:T4-2)

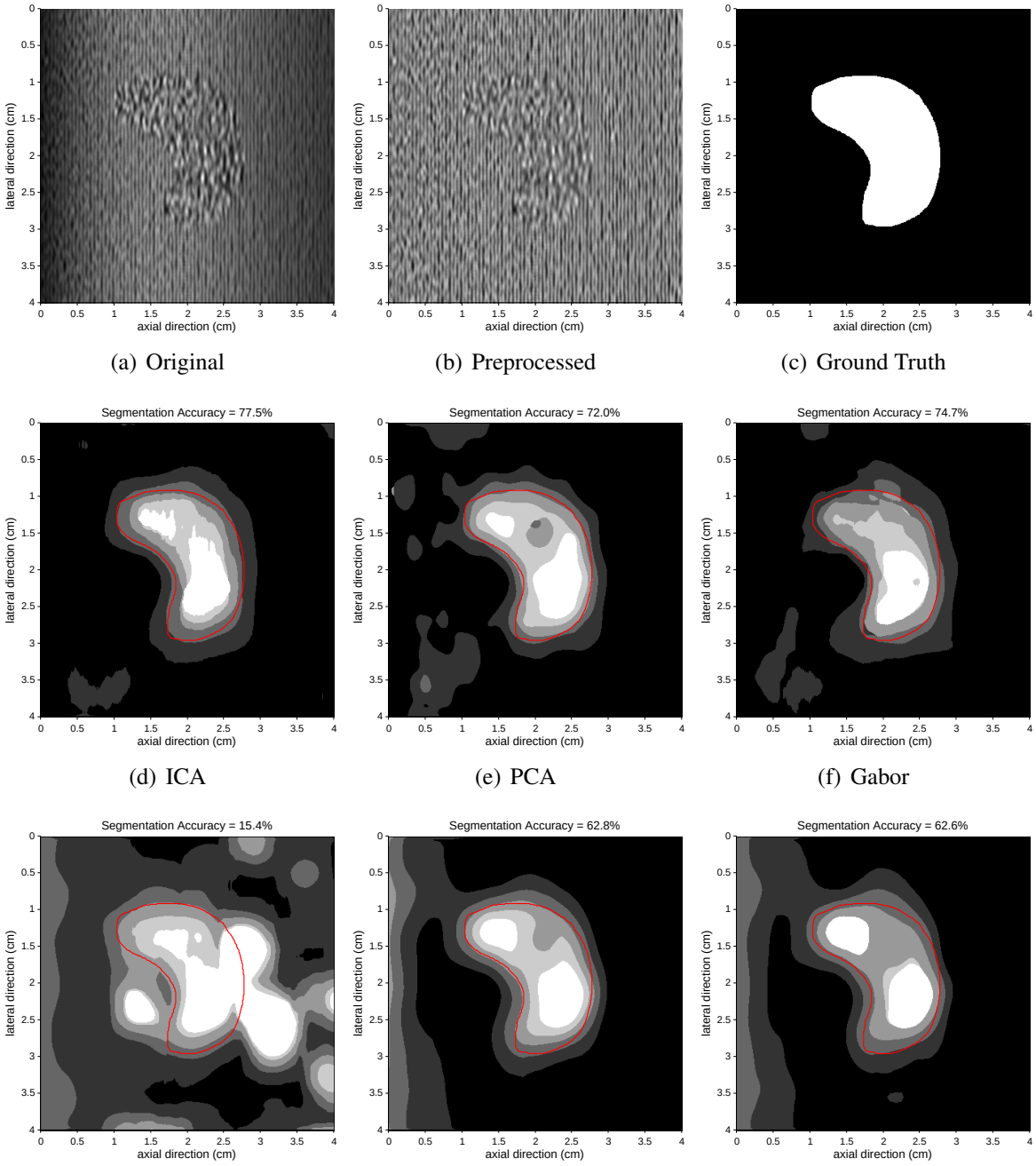


(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.55: Two-component Texture Segmentation Results (Background:T4-3, Lesion:T4-5)



(a) Original (b) Preprocessed (c) Ground Truth

(d) ICA (e) PCA (f) Gabor

(g) FWHM (h) SNRALPHA (i) SNRGAMMA

Figure E.56: Two-component Texture Segmentation Results (Background:T4-0, Lesion:T4-5)

Bibliography

[1] D. Blanco and B. Mulgrew. ICA in signals with multiplicative noise. *IEEE Transaction on Signal Processing*, 53(8):2648–2657, August 2005.

[2] Nizar Bouhlel, Sylvie Sevestre, Hatem Rajhi, and Radhi Hamza. A new markov random field model based on K-distribution for textured ultrasound image. In *Medical Imaging: Ultrasonic Imaging and Signal Processing, Proc. SPIE 5373*, pages 363–372, April 2004.

[3] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Trans. Pattern Analysis and Machine Intelligence.*, 12:55–73, January 1990.

[4] C.B. Burckhardt. Speckle in ultrasound B-mode scans. *IEEE Trans. Sonics Ultrason.*, 25(1):273–278, 1978.

[5] L. Busse, T.R. Crimmins, and J.R. Fienup. A model based approach to improve the performance of the geometric filtering speckle reduction algorithm. In *Proc. IEEE Ultrason. Symp.*, pages 1353–1356, 1995.

[6] T. Chang and C.-C. J. Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE Trans. Image Processing*, 2(4):429–441, 1993.

[7] C. H. Chen and Xianju Wang. Ultrasonic NDE image enhancement with nonlinear filters in signal subspaces. In *16th WCNDT 2004 - World Conference on NDT*, September 2004.

[8] Yen-Wei Chen, Xiang-Yan Zeng, and Hanqing Lu. Edge detection and texture segmentation based on independent component analysis. In *16th International Conference on Pattern Recognition (ICPR'02)*, volume 3, pages 351–354, 2002.

[9] Yunqiang Chen, Hongcheng Wang, Tong Fang, and Jason Tyan. Image compounding based on independent noise constraint. In *Proc. IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*, March 2005.

[10] Jian Cheng, Qingshan Liu, and Hanqing Lu. Texture classification using kernal independent component analysis. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, 2004.

[11] Andrzej Cichoki and Shun ichi Amari. *Adaptive Blind Signal and Image Processing, Learning Algorithms and Applications*. John Wiley & Sons, LTD, 2002.

[12] Robert L. Cook. Stochastic sampling in computer graphics. *ACM Trans. Graph.*, 5(1):51C72, 1986.

[13] Nello Cristianini and John Shawe-Taylor. *An introduction to Support Vector Machines and other kernel-based learning methods*. Cambrige University Press, 2000.

[14] David L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.

[15] V. Dutt. *Statistical analysis of ultrasound echo envelope*. PhD thesis, Mayo graduate school, Rochester, MN, 1995.

[16] V. Dutt and J.F. Greenleaf. Statistics of log-compressed echo envelope. *J. Acoust. Soc. Am.*, 99:3817–3825, 1996.

[17] Simone Fiori. Overview of independent component analysis techniques with an application to synthetic aperture radar(SAR) imagery processing. *Neural Networks archive 2003 Special issue: Neural network analysis of complex scientific data: Astronomy and geosciences*, pages 453–467, 2003.

[18] V.S. Frost, J.A. Stiles, K.S. Shanmuggam, and J.C. Holtzman. A model for radar images and its application for adaptive digital filtering of multiplicative noise. *IEEE Trans. Pattern Anal. Machine Intell.*, 4(2):157–165, 1982.

[19] L. Gagnon and A. Jouan. Speckle filtering of SAR images - A comparative study between complex-wavelet based and standard filters. *SPIE*, 3169:80–91, 1997.

[20] Joshua Gluckman. Visually distinct patterns with matching subband statistics. *IEEE Trans. Pattern Analysis and Mach. Intelligence*, 27(2):252–264, February 2005.

[21] Osman Gokhan, Aysin Ertuzun, and Aytul Ercil. Independent component analysis for texture defect detection. *Pattern Recognition and Image Analysis*, 14-2:303–307, 2004.

[22] S. Gupta, R. C. Chauhan, and S. C. Sexana. Wavelet-based statistical approach for speckle reduction in medical ultrasound images. *Med. Biol. Eng. Comput.*, 42:189–192, 2004.

[23] R. M. Haralick, K. Shanmugam, and I. Dinastein. Textural features for image classification. *IEEE Trans. Syst. Man Cyber.*, 3:614–624, 1973.

[24] Michel Haritopoulos, Hujun Yin, and Nigel M. Allinson. Multiplicative noise removal using self-organizing maps. In *Proc. of ICA 2001*, pages 206–211, 2001.

[25] Lennart Hellström. *Hilbert Transforms in Signal Processing*. Artech House, Inc., Boston, 1996.

[26] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. *A Practical Guide to Support Vector Classification*. Department of Computer Science, National Taiwan University, Taipei 106, Taiwan, March 2008.

[27] Aapo Hyvärinan, Patrik Hoyer, and Erkki Oja. Sparse coding shrinkage: Denoising by nonlinear maximum likelihood estimation. In *Proceedings of the 1998 conference on Advances in neural information processing systems*, pages 473–479, 1999.

[28] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[29] Aapo Hyvärinen. Sparse code shrinkage: denoising of nongaussian data by maximum likelihood estimation. *IEEE Signal Processing Letters*, 17(7):1739–1768, 1999.

[30] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. John Wiley & Sons, Inc., 2001.

[31] Yellot J. I. Spectral analysis of spatial sampling by photoreceptors: Topological disorder prevents aliasing. *Vision Research*, 22:1205C1210, 1982.

[32] Yellot J. I. Spectral consequences of photoreceptor sampling in the rhesus retina. *Science*, 221:382C385, 1983.

[33] M.F. Insana, R.F. Wagner, B.S. Garra, D.G. Brown, and T.H. Shawker. Analysis of ultrasound image texture via generalized rician statistics. *Opt. Eng.*, 25:743–748, 1986.

[34] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[35] E. Jakeman. Speckle statistics with a small number of scatterers. *Opt. Eng.*, 23:453–461, 1984.

[36] E. Jakeman and R.J.A. Tough. Generalized K-distribution: A statistical model for weak scattering. *J. Opt. Soc. Am.*, 4:1764–1772, 1987.

[37] J.A. Jensen. Field: A program for simulating ultrasound systems. In *10th Nordic-Baltic Conference on Biomedical Imaging Published in Medical & Biological Engineering & Computing*, volume 34-1, pages 351–353, 1996.

[38] J.A. Jensen. Linear description of ultrasound imaging systems - Notes for summer school on advanced ultrasound imaging. Technical report, 1999.

[39] S. Jin, Y. Wang, and J. Hiller. An adaptive non-linear diffusion algorithms for filtering medical images. *IEEE Trans. Inform. Technol. Biomed.*, 4(4):298–305, December 2000.

[40] I.T. Jolliffe. *Principal Component Analysis*. Springer, 2 edition, 2002.

[41] B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frish. Inability of humans to discriminate between visual textures that agree in second-order statistics - revisited. *Perception*, 2:391–405, 1973.

[42] B. Julesz, E. N. Gilbert, and J. D. Victor. Visual discrimination of textures with identical third-order statistics. *Biol. Cybernetics*, 31:137–140, 1978.

[43] Sam T. Kaplan and Tadeusz J. Ulrych. Blind deconvolution and ICA with banded mixing matrix. In *4th Intl. Symp. on ICA and BSS (ICA 2003)*, Nara, Japan, April 2003.

[44] Juha Karvonen and Markku Simila. Independent component analysis for sea ice SAR image classfication. *Geoscience and Remote Sensing Symposium, 2001. IGARSS '01. IEEE 2001 International*, 3:1255–1257, 2001.

[45] C. Kotropoulos and I. Pitas. Optimum nonlinear signal detection and estimation in the presence of ultrasonic speckle. *Ultrason. Imag.*, 14:249–275, 1992.

[46] Constantine Kotropoulos and Ioannis Pitas. Segmentation of ultrasonic images using support vector machines. *Pattern Recognition Letters*, 24(2003):715–727, 2003.

[47] Karl Krissian, Kirby Vosburgh, Ron Kikinis, and Carl-Fredrik Westin. Anisotropic diffusion of ultrasound constrained by speckle noise model. Technical report, Harvard Medical School Brigham and Women's Hospital Department of Radiology, 75 Francis Street 02115 Boston, MA, USA, October 2004.

[48] Ares Lagae and Philip Dutré. A comparison of methods for generating poisson disk distributions. *Computer Graphics Forum*, 2007.

[49] Te-Won Lee and Michael S. Lewichi. Unsupervised image classification, segmentation and enhancement using ICA mixture models. *IEEE trans. image process.*, 11-3:270–279, 2002.

[50] Wen-Li Lee and Yung-Chang Chen. Ultrasonic liver tissue classfication by fractal feature vector based on m-band wavelet transform. *IEEE Trans. On Med. Imaging*, 22(3), March 2003.

[51] Yi-Ou Li, Tulay Adah, and Vince D. Calhoun. Independent component analysis with feature selective filtering. In *Proceedings of the 2004 14th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, pages 193–202, 2004.

[52] Xiuwen Liu and Lei Cheng. Independent filters for texture classification. *IEEE ICIP 2002*, III:113, 2002.

[53] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[54] Roberto Manduchi and Javier Portilla. Independent component analysis of textures. In *Proceedings of the 7th IEEE International Conference on Computer Vision*, 1999.

[55] Jean-Christophe Pesquet Marc Castella. An iterative blind source separation method for convolutive mixture of images. *Independent Component Analysis and Blind Signal Separation*, 3195/2004:922–929, October 2004.

[56] P. Mayo, Francisco Rodenas Escriba, and Gumersindo Verdu Martin. Denoising mammographic images using ICA. *ICA 2004, LNCS 3195*, pages 1064–1071, 2004.

[57] F.N.S. Medeiros, N.D.A. Mascarenhas, R.C.P. Marques, and C.M. Laprano. Edge preserving wavelet speckle filering. In *Proc. 5th IEEE Southwest Symp. Image Anal. Interpretation*, pages 281–285, April 2002.

[58] O. V. Michailovich and D. Adam. Robust estimation of ultrasound pulses using outlier-resistant de-noising. *IEEE Transcations on Medical Imaging*, 22:368–392, March 2003.

[59] O. V. Michailovich and A. Tannenbaum. Despeckling of medical ultrasound images. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 53(1):64–78, January 2006.

[60] Thanassis Misaridis. *Ultrasound Imaging Using Coded Signals*. PhD thesis, Center for Fast Ultrasound Imaging, Technical University of DenmarkMayo graduate school, Denmark, 2001.

[61] Aleksandra Mojsilovic, Miodrag V. Popovic, Aleksandar N. Neskovic, and Aleksandar D. Popovic. Wavelet image extsion for analysis and classification of infarcted myocardial tissue. *IEEE Trans. On Biomed. Eng.*, 44(9), September 1997.

[62] Radu Muthihac and Marc M. Van Hulle. Statistics of feature extraction by topographic independent component analysis from natural images. In *Proc. of the 2002 WSEAS Int'l. Conf on Electronics, Control and Signal Processing*, pages CD–ROM paper 451–289, Singapore, December 2002.

[63] B.J. Oosterveld, J.M. Thijssen, and W.A. Verhoef. Texture in B-mode echograms: 3-d simulations and experiments of the effects of diffraction and scatterer density. *Ultrason. Imag.*, 7:142–160, 1985.

[64] D. Pollen and S. Ronner. Visual cortical neurons as localized spatial frequency filters. *IEEE Trans. on Systems, Man, and Cybernetics*, 13:907–916.

[65] William K. Pratt. *Digital Image Processing*. John Wiley & Sons, Inc., 2007.

[66] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes - The Art of Scientific Computing*. Cambridge University Press, 2007.

[67] Trygve Randen and John Hakon Husoy. Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):291–310, April 1999.

[68] Navalgund Rao. Ultrasound imaging. In J. Hornak, editor, *Encyclopedia of Imaging Science & Technology*, pages 1412–1435. John Wiley & Sons, 2002.

[69] J. Saniie, T. Wang, and N. Bilgutay. Analysis of homomorphic processing for ultrasonic grain signal characterization. *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, 3:365–375, 1989.

[70] A. Serdaroglu, A.B.Ertuzun, and A.Ercil. Defect detection in textile fabric images using wavelet transforms and independent component analysis. *Pattern Recognition and Image Analysis*, 16:61–64, January 2006.

[71] P. M. Shankar, J. M. Reid, H. Ortega, C. W. Piccoli, and B. B. Goldberg. Use of non-rayleigh statistics for the identification of tumors in ultrasonic b-scans of the breast. *IEEE Trans. Med. Imag.*, 12(4):687–692, 1993.

[72] P.M. Shankar, V.A. Dumane, J.M. Reid, V. Genia, T. George, F. Forsberg, C.W. Piccoli, and B.B. Goldberg. Use of k-distribition for classification of breast masses. *Ultrasound Med. Biol.*, 26(9):1503–1510, 2000.

[73] Mark A. Sheppard and Liwen Shih. Efficient image texture analysis and classification for prostate ultrasound diagnosis. In *Proceedings of the 2005 IEEE Computational Systems Bioinformatics Conference Workshops (CSBW05)*, 2005.

[74] E. P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representation. *Annu. Rev. Neurosci*, 24:193–216, 2001.

[75] Mohamed S.S., Abdef-Galil T.K., Salama M.M.A., Fenster A., Downey D.B., Rizkalla K., El-Saadany E.F., and Kamel M. Prostate cancer diagnosis based on gabor filter texture segmentation of ultrasound image. In *Proceedings of CCECE 2003, Montreal*, pages 1485–1488, May 2003.

[76] P. R. Stepanishen. The time-dependent force and radiation impedance on a piston in a rigid infinite planar baffle. *J. Acoust. Soc. Am.*, 49:841–849, 1971.

[77] P. R. Stepanishen. Transient radiation from pistons in an infinite planar baffle. *J. Acoust. Soc. Am.*, 49:1629–1638, 1971.

[78] Thomas L. Szabo. *Diganostic Ultrasound Imaging: Inside Out*. Elsevier Academic Press, 2004.

[79] Johan M. Thijssen. Ultrasonic speckle formation, analysis and processing applied to tissue characterization. *Pattern Recognition Letters*, 24:659–675, 2003.

[80] G. E. Tupholme. Generation of acoustic pulses by baffled plane pistons. *Mathematika*, 16:209–224, 1969.

[81] M. Tur, K. C. Chin, and J. W. Goodman. When is speckle noise multiplicative? *Applied Optics*, 21(7):1157–1159, 1982.

[82] F.M.J Valckx and J.M. Thijssen. Characterization of echographic image texture by co-occurrence matrix parameters. *Ultrasound Med. Biol.*, 23:559–571, 1997.

[83] J. H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Royal. Society ser. B*, 265:359–366, 1998.

[84] J.T.M. Verhoeven and J.M. Thijssen. Improvement of lesion detectability by speckle reduction filtering: A quantitative study. *Ultrason. Img.*, 15:181–204, 1993.

[85] R.F. Wagner, M.F. Insana, and D.G. Brown. Unified approach to the detection and classification of speckle texture in diagnostic ultrasound. *Opt. Eng.*, 25:738–742, 1986.

[86] R.F. Wagner, S.W. Smith, J.M. Sandrik, and H. Lopez. Statistics of speckle in ultrasound B-scans. *IEEE Trans. Sonics Ultrason.*, 30(7):156–163, 1983.

[87] T. P. Weldon and W. E. Higgins. Design of multiple gabor filters for texture segmentation. In *Proc. Int'l Conf. Acoustic Speech, Signal Proc.*, pages 2243–2246, Atlanta, GA, May 1996.

[88] Chung-Ming Wu, Yung-Chang Chen, and Kai-Sheng Hsieh. Texture features for classification of ultrasonic liver images. *IEEE Trans. Med. Imag.*, 11(2):141–152, June 1992.

[89] Xiang-Yan Zeng, Yen-Wei Chen, and Zensho Nakao. Image feature representation by the subspace of nonlinear pca. In *16th International Conference on Pattern Recognition (ICPR'02)*, volume 2, pages 228–231, 2002.

[90] Xiaohui Zhang and Chi Hau Chen. New independent component analysis method using higher order statistics with application to remote sensing images. *Opt. Eng.*, 41(7):1717–1728, July 2002.